THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Quantitative Cone-Beam Computed Tomography Reconstruction for Radiotherapy Planning

Jonathan Hugh Mason

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
October 24, 2018

Abstract

Radiotherapy planning involves the calculation of dose deposition throughout the patient, based upon quantitative electron density images from computed tomography (CT) scans taken before treatment. Cone beam CT (CBCT), consisting of a point source and flat panel detector, is often built onto radiotherapy delivery machines and used during a treatment session to ensure alignment of the patient to the plan. If the plan could be recalculated throughout the course of treatment, then margins of uncertainty and toxicity to healthy tissues could be reduced. CBCT reconstructions are normally too poor to be used as the basis of planning however, due to their insufficient sampling, beam hardening and high level of scatter. In this work, we investigate reconstruction techniques to enable dose calculation from CBCT. Firstly, we develop an iterative method for directly inferring electron density from the raw X-ray measurements, which is robust to both low doses and polyenergetic artefacts from hard bone and metallic implants. Secondly, we supplement this with a fast integrated scatter model, also able to take into account the polyenergetic nature of the diagnostic X-ray source. Finally, we demonstrate the ability to provide accurate dose calculation using our methodology from numerical and physical experiments. Not only does this unlock the capability to perform CBCT radiotherapy planning, offering more targeted and less toxic treatment, but the developed techniques are also applicable and beneficial for many other CT applications.

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the School of Engineering at The University of Edinburgh.

Jonathan Hugh Mason

for Nick

Acknowledgements

I would like to firstly wholeheartedly thank my supervisor Prof. Mike Davies. He gave me the opportunity to start this journey, and has helped me tremendously throughout. He had a very active role in this project, and inspired many of its developments, while giving me the freedom to explore and take the work in new directions. I have learnt a great deal from him, and he is an academic I truly admire. It has been an absolute privilege to work with him.

My second supervisor, Dr. Bill Nailon, has also been instrumental in this project. He has been an great source of knowledge on radiotherapy, has acquired data in his evenings on the clinical scanners, and has made significant contributions to every piece of work we produced. Dr. Alessandro Perelli was also a collaborator on this project throughout, and I thank him for all the productive discussions, and his many timely and useful reviews. It has been a pleasure to work with them both.

I also give thanks to my examiners Prof. William Lionheart and Dr. Nick Polydorides for their time evaluating the manuscript, and their useful feedback and critique, which has allowed me to significantly improve the quality of this thesis.

A key source of help, that really unlocked the potential of this project, was assistance from Dr. Adam Wang at Varian Medical Systems. Not only did he pre-process our measurements for us, but he offered useful information and advice about the system, without which we would not have been able to exploit the clinical systems at the Western General Hospital. I would also like to sincerely thank Prof. Jeff Siewerdsen for establishing this connection, and also providing invaluable knowledge and insight into cone-beam imaging.

It took a while before I had work that was positively reviewed. The comments from the four rejected manuscripts in the first two years helped me steer the project into productivity, and made me really raise the standards of both the science and its presentation. I am truly grateful to all the anonymous reviewers that have taken the time to evaluate our work, especially those who provided such productive criticism.

In November 2017, I had the opportunity to spend three weeks at medPhoton in Salzburg, Austria, which was an enlightening experience that really expanded my understanding of the cone-beam system. I am truly thankful to Heinz Deutschmann for making this happen and hosting me in his company, and to Dr. Daniel Kellner for his extensive help before and throughout. I would also like to thank everyone else at the company too, for assisting me with my work, openly discussing their products, and making me feel very welcome.

My affinity for CT reconstruction started during my Master's thesis project in the TOMCAT group at the Paul Sherrer Institut, Switzerland. I thank Prof. Marco Stampanoni and Dr. Federica Marone for giving me this opportunity. I am also indebted to Dr. Filippo Arcadu for introducing me to academic research, and providing such great guidance and supervision during my project.

I have also had incredible personal support throughout this process. I thank Nina for her love and support through all the highs and lows, and for making these years so enjoyable — she is the most wonderful person to share life with.

I am eternally grateful to my parents for everything they have given me, and to Jenny for being such a great sister. My brother Nick has been and continues to be a huge inspiration. He was a brilliant man and scientist, and I aspire to his curiosity and creativity.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ADC	analogue to digital converter
ART	algebraic reconstruction technique
CBCT	cone-beam computed tomography
CT	computed tomography
fASKS	fast adaptive scatter kernel superposition [8]
FBP	filtered backprojection
FDK	Feldkamp–Davis–Kress method [9]
FFT	fast Fourier transform
FISTA	fast iterative shrinkage-thresholding algorithm [10]
IFFT	inverse fast Fourier transform
IMPACT	iterative maximum-likelihood polychromatic algorithm for CT [11]
NLL	negative log-likelihood
OAR	organs at risk
PICCS	prior image constrained compressed sensing [12]
Polyquant	polyenergetic quantitative reconstruction method
PolySKS	polyenergetic scatter kernel superposition
PTV	planning target volume
PWLS	penalised weighted least squares
RMSE	root mean squared error
SART	simultaneous algebraic reconstruction technique
SIRT	simultaneous iterative reconstruction technique
SKS	scatter kernel superposition
TV	total variation
WLS	weighted least squares

Nomenclature

N_{proj}	number of projections
N_{ray}	number of measurements
N_{vox}	number of discrete volumetric elements (voxels)
N_{ξ}	number of discrete spectrum samples
m	mass attenuation coefficient (cm^2/g)
μ	linear attenuation coefficient (cm^{-1})
ρ	mass density (g/cm^3)
$\boldsymbol{\mu} \in \mathbb{R}_+^{N_{\text{vox}}}$	attenuation coefficient of specimen
$\boldsymbol{\rho}_e \in \mathbb{R}_+^{N_{\text{vox}}}$	relative electron density of specimen
$\boldsymbol{p} \in \mathbb{R}_+^{N_{\text{ray}}}$	linearised projections ('post-log')
$\boldsymbol{y} \in \mathbb{R}_+^{N_{\text{ray}}}$	raw X-ray measurements ('pre-log')
$\boldsymbol{b} \in \mathbb{R}_+^{N_{\text{ray}}}$	incident (flat field) intensity
$\boldsymbol{s} \in \mathbb{R}_+^{N_{\text{ray}}}$	scatter intensity signal
$\boldsymbol{\Phi} \in \mathbb{R}_+^{N_{\text{ray}} \times N_{\text{vox}}}$	system matrix and forward-projection operator
$\boldsymbol{\Phi}^T \in \mathbb{R}_+^{N_{\text{vox}} \times N_{\text{ray}}}$	backprojection operator

Chapter 1

Introduction

1.1 Project Motivation and Scope

There are over 350,000 new cases of cancer each year in the UK alone [13], of which 50% are fatal. Radiotherapy is the use of high energy electromagnetic radiation or particles to irradiate the cancerous tissue, with the aim of destroying it or impeding its growth, and is currently used in 27% of cancer cases in the UK [14]. If the efficacy of radiotherapy could be enhanced, then its success rate and the mitigation of complications is likely to be improved.

Radiotherapy planning involves designing the delivery of radiation to have maximum effect on the cancerous tissue, whilst sparing healthy organs that are at risk of becoming damaged or may develop secondary cancer from the ionising beam. The delivery itself is commonly fractionated over as many as 50 days, which gives healthy tissue the chance to recover from the effects of radiation [15]. In order to inform this planning procedure, medical imaging is used to get structural and physical information about the patient's body. In particular, X-ray computed tomography (CT) is an integral part of this process.

CT is the process of gathering sets of X-ray transmission images of a specimen of interest, and combining them to infer its spatial material properties. CT imaging is very complementary to the planning process, since the mechanisms of image contrast are the same as the tissue's interaction with the radiotherapy beam. With this, it can used to calculate the dose deposition and allows the optimisation of the plans. Therefore, CT imaging offers both structural information about organ size, shape and location, and a quantitative map for the purpose of planning.

The quantity that is most prevalent in this process is the electron density, as it is electrons that account for the interaction and energy absorption of the treatment beam, especially in the case of high energy electromagnetic radiation — also known as photon therapy. The

mapping from a collection of X-ray measurements to electron density is most commonly performed in three distinct steps. First is the linearisation of the measurements, involving the correction of effects such as non-linearity from the spectrally diverse source and X-ray scatter contamination, and conversion to a projection of attenuation. Next, these linear projections are inverted to form a 3D map of X-ray attenuation, known as reconstruction. Finally, the reconstructed images are non-linearly converted to electron density through a calibration process. The treatment calculation itself is often performed only once at the start of treatment, with the use of an enclosed gantry fan-beam CT, which is regarded as a fast and accurate CT geometry. This means that the time between the planning and delivery will be between days and several weeks throughout the course of treatment, and the inevitable movement and changes in organ shape and position must be accounted for by adding significant margins of error around the treatment target. Therefore, these margins form part of the treatment prescription [16, 17], though at the expense of the surrounding healthy tissue.

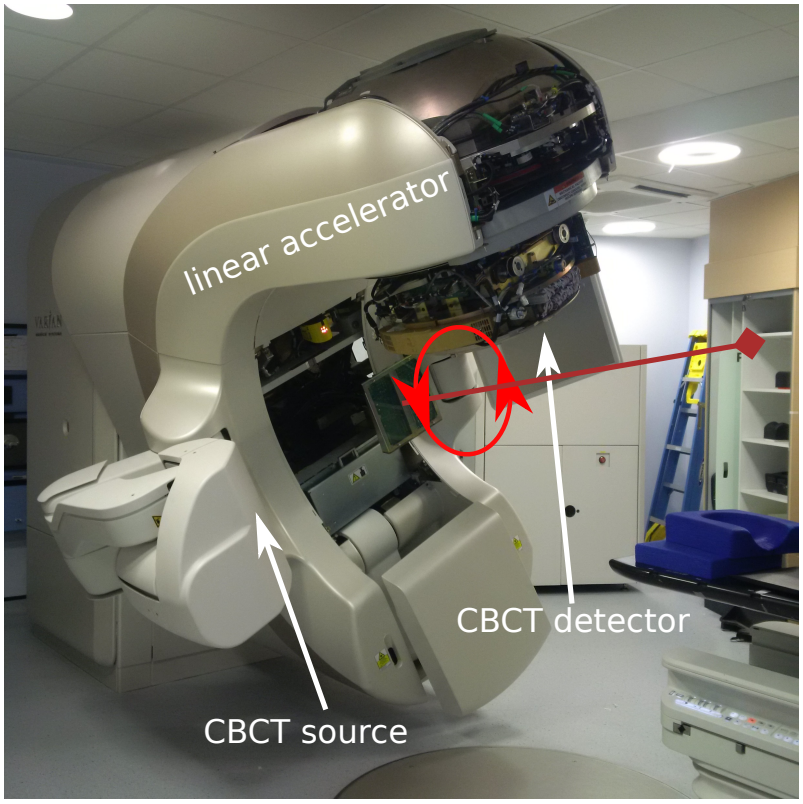


Figure 1.1: Photograph of photon therapy unit with CBCT imaging system. The line in dark red indicates the axis of gantry rotation, with an anticlockwise rotation shown with the light red arrow.

The cone-beam CT (CBCT) geometry consists of a divergent source and planar detector,

which are often circularly rotated around the patient, although any smooth movement of the source is also possible. CBCT has extensive use during radiotherapy, since it has a large field of view, has the potential to be portable, or can be directly mounted onto the treatment hardware. An example of a CBCT system mounted onto a radiotherapy device is shown in Figure 1.1. Despite its popularity, it also suffers from several fundamental limitations. Firstly, the circular scan that is most commonly used, has a sampling deficit away from the centre of rotation [18], that means that fully accurate imaging is not possible without additional constraints. There are also prevalent physical artefacts: non-linearity of the X-ray attenuation due to polyenergetic source and energy indiscriminant detectors; and large amounts of object scatter contamination due to the large field of view. On top of this, there are mechanical difficulties due to the movement of the heavy imaging components through rotation.

Due to these limitations in CBCT, although systems are often employed for tracking structural changes and aligning a planning CT from an enclosed gantry fan-beam system, they are not currently used for the basis of treatment dose calculation [19–21]. If their quality could be sufficiently enhanced however, then this may offer large clinical advantages. As the plan could be updated to more accurately capture the current anatomical structure, the margins of error around the target could be reduced, which will reduce the toxicity to the surrounding healthy tissues. With the potential for more targeted treatment, the plans may also be modified to deliver higher doses over fewer fractions, and could ultimately enable a higher throughput of patients for a given radiotherapy resource. In order to realise this potential, a technological gap must be closed, which is the ability to perform quantitatively accurate imaging with CBCT, ideally making it comparable to traditional enclosed gantry fan-beam CT. This is the goal of this project.

In this work, we will attempt to correct for the most significant deficits in CBCT, to improve its quantitative accuracy. At the core of this, we will employ statistical reconstruction with sparsity promoting regularisation to compensate for the sampling deficit. Into this, we will develop a model for accounting for the non-linearity of the attenuation from a polyenergetic source, and actively model the scatter contamination. Ultimately, we wish to have a model that is able to take a set of 'raw' contaminated X-ray measurements from CBCT, and directly infer the spatial quantitative electron density of the specimen, to an accuracy adequate for dose calculation. Ideally, the method would facilitate on-line adaptive planning whilst the patient is lying on the couch, so speed

is also a key consideration, and we must be conscious of the computational cost of our methodology.

1.2 Main Contributions

This thesis proposes a methodology for direct and accurate quantitative CBCT reconstruction. We identify the main contributions of this work as:

1. Introducing a method for modelling the polyenergetic attenuation of biological tissues and metallic implants in terms of physical quantities, and using this as the basis for the direct inference of electron density through statistical reconstruction — we have coined this method Polyquant.

2. Developing iterative algorithms exploiting the Polyquant model, able to incorporate sparse regularisation functions. This includes a forward-backward splitting algorithm inspired by the iPiano method for nonconvex objective functions [22], and a greatly accelerated variant employing 'ordered subsets with momentum' of [23].

3. Introducing a new convolutional scatter model, dependent on the polyenergetic and electron density information available at each iteration, which we denote PolySKS. We then fuse this model into an iterative Polyquant reconstruction, denoted Polyquant–PolySKS, which provides adaptive scatter compensation with minimal additional computational cost.

4. Demonstration through numerical ground truth experiments and real phantom data that these methods are able to provide superior electron density estimation over competing approaches.

5. Through both diverse numerical experiments from several different anatomical sites and radiation types, and against physical measurements from a real clinical radiotherapy machine, we show the enhancement in dose calculation accuracy with these methods. In the real data case, we demonstrate we are not only able to greatly outperform the accuracy from the commercial CBCT techniques, but also images from enclosed gantry fan-beam CT — generated using traditional reconstruction.

1.2.1 Project Outputs

Journal Articles

1. **Quantitative cone-beam CT reconstruction with polyenergetic scatter model fusion**
   Jonathan H Mason, Alessandro Perelli, William H Nailon, Mike E Davies
   Physics in Medical and Biology, [accepted], 2018

2. **Polyquant CT: direct electron and mass density reconstruction from a single polyenergetic source**
   Jonathan H Mason, Alessandro Perelli, William H Nailon, Mike E Davies
   Physics in Medical and Biology, **62** 8739, 2017

Conference Proceedings

1. **A Monte Carlo framework for low dose CT reconstruction testing**
   Jonathan H Mason, William H Nailon, Mike E Davies
   International Workshop on Simulation and Synthesis in Medical Imaging, 2017

2. **Quantitative electron density CT imaging for radiotherapy planning***
   Jonathan H Mason, Alessandro Perelli, William H Nailon, Mike E Davies
   Annual Conference on Medical Image Understanding and Analysis, 2017
   * awarded prize for best paper of conference

3. **Can planning images reduce scatter in follow-up cone-beam CT?**
   Jonathan H Mason, Alessandro Perelli, William H Nailon, Mike E Davies
   Annual Conference on Medical Image Understanding and Analysis, 2017

4. **Framework for statistical cone-beam CT reconstruction with prior Monte Carlo scatter estimation**
   Jonathan H Mason, William H Nailon, Mike E Davies
   ESTRO 36, 2017

5. **Exploiting planning CT data for accurate WEPL on CBCT reconstructions used in adaptive radiotherapy**
   Jonathan H Mason, William H Nailon, Mike E Davies

ESTRO 35, 2016

1.3 Thesis Organisation

This thesis begins with background material in Chapter 2. This includes: an overview of the role of imaging in radiotherapy planning; fundamental information on the physics relevant for CT imaging; a description of the CBCT system hardware; the mathematics of CT imaging, with a review of reconstruction techniques and artefact correction approaches; and finally, a description of the numerical simulation of CT systems as is used extensively throughout this work.

In Chapter 3, we present the Polyquant method for direct quantitative inference from a polyenergetic source. Since this concept transcends CBCT and electron density imaging, we also actively evaluate other applications such as a fan-beam CT geometries and mass density imaging. In this chapter, we introduce a statistical iterative reconstruction algorithm and its accelerated variant. This approach is justified through extensive evaluations of the model, numerical ground truth reconstructions, and with real CBCT data.

Chapter 4 then sees the introduction of our polyenergetic scatter model, which we develop with reference to an alternative approach to estimating scatter during preprocessing. We provide details on the integration of the model into Polyquant, giving the proposed Polyquant–PolySKS method. As in Chapter 3, this is then tested through numerical and real data experiments.

We then put the developed methods to the test for our application in Chapter 5, where we evaluate their accuracy in dose calculation against a host of competing techniques, and in a range of different radiotherapy settings.

Finally, we provide conclusions in Chapter 6, along with an extensive discussion on future work and the applicability of the developed methods within and outside of radiotherapy. Since this work is designed eventually with medical use in mind, we will give special attention to the gap between the developed techniques and their possible use in a clinical setting.

1.4 Notation

Due to the various types of functions and variables used throughout this thesis, a clear and consistent notation convention is important.

Firstly, we will denote scalar variables in normal weight and discrete vectors in bold weight. For example, we could define

$$\boldsymbol{a} \equiv \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \tag{1.1}$$

for the column vector $\boldsymbol{a} \in \mathbb{R}^n$. As is used in (1.1), subscripts indicate the index within the vector. The text weighting also applies for functions, where $f(a)$ and $f(\boldsymbol{a})$ both return scalars, and $\boldsymbol{f}(a)$ and $\boldsymbol{f}(\boldsymbol{a})$ return discrete vectors. Where possible, matrices will be denoted in both upper case and bold weight such as $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. The elements within a matrix are also indexed with subscripts such as $A_{2,4}$ being the 4^{th} column of the 2^{nd} row of $\boldsymbol{A}$.

For algebraic convenience, all discrete variables will be column vectors as (1.1). Since the application is imaging however, these vectors will have important spatial arrangement, so extra notation is useful for describing their organisation in space. The two spaces of relevance in CT imaging are volumetric space and measurement space.

Measurement space for CBCT is a stack of 2D images, which we denote as a column vector such as $\boldsymbol{p} \in \mathbb{R}^{N_{\text{ray}}}$, where N_{ray} is the number of measurements. We will also use the spatially meaningful indexing $\boldsymbol{p}\{i\}$, where i is the index of a projection, and selects all indexes of the column vector corresponding to that projection. On top of this, where indexing within a given projection itself is appropriate, we use the notation $p\{i\}(u, v)$, where (u, v) are the 2D Cartesian coordinates of the detector plane for the i^{th} projection.

Discrete volumetric space is defined in 3D, though again we will algebraically represent this with column vectors $\boldsymbol{v} \in \mathbb{R}^{N_{\text{vox}}}$, with N_{vox} the number of volumetric elements or 'voxels'. Where appropriate, we will refer to a function at a physical location as $v(x, y, z)$.

Wherever a product between vectors or matrices is shown such as $\boldsymbol{b}^T\boldsymbol{a}$ or $\boldsymbol{A}\boldsymbol{b}$, then this represents the matrix product, where $\cdot^T$ is the transpose operation. Element-wise multiplication is denoted with $\odot$ such as $\boldsymbol{A} \odot \boldsymbol{B}$, and the symbol $\oslash$ is used for element-wise division.

In some areas, we will refer to the continuous space rather than discrete, particularly in the volumetric space. Here, we are only dealing with scalar fields, which we will denote in regular weight, and will use either $g(\boldsymbol{a})$ or $g(x, y, z)$ as appropriate. We will strive to make the distinction between continuous and discrete variables clear.

One final comment on notation is in the use of superscripts, which will serve two roles. One of which is to represent powers, for example $a^2 \equiv aa$. Powers are also always represented as element-wise operations, so $\boldsymbol{a}^2 \equiv \boldsymbol{a} \odot \boldsymbol{a}$. The other use of superscripts is for the indexing of iterative methods, such as

$$a^{k+1} := a^k + 1,$$

which increments the scalar a by 1, and the symbol $:=$ is used to show this is an iterative update. Since this notation is overloaded, there are cases where one must be careful to avoid ambiguity. Wherever both a power and iterative index are used in conjunction, we will use parentheses, where for example

$$a^{k+1} := (a^k)^2,$$

iteratively squares the scalar a. The other potential problem may be if one wanted to use the iterative index k to raise a variable to a power. Since we will never do this however, this can be ignored for this work.

Chapter 2

Background

In this chapter, we will present an overview on CBCT imaging for radiotherapy. This will begin with the role of general CT imaging in radiotherapy in Section 2.1, with a typical work-flow and an introduction to the dose calculation procedure. We also provide a description of various forms of radiation used in cancer control, which we will later use to evaluate our developed methodology in Chapter 5.

Next, we will present the physics relevant for digital X-ray imaging in Section 2.2: describing the mechanisms of generation, attenuation within the patient, and detection through scintillation and photo-detection. These will then be built upon in our description of CT and CBCT systems in Section 2.3, with definitions of the measurement and noise properties that will be used extensively in Chapter 3 and Chapter 4.

The review on CT reconstruction then comes in two parts: Section 2.4 describes fundamentals of image formation, including classical methods such as filtered backprojection and algebraic iterative reconstruction, and discretisation schemes; Section 2.5 then establishes the concept of model based reconstruction, and reviews recent developments in this area. The methods we later develop can themselves be considered model based iterative algorithms, so we will build upon the work presented here, especially in Chapter 3, and use the reviewed techniques as bases of comparison.

Section 2.6 provides a review of simple beam hardening and scatter correction approaches, that we will build upon in Chapter 3 and Chapter 4 respectively.

Finally, we provide a description of the simulation tool we use extensively for numerical ground truth experiments in Section 2.7. This consists of a suite of freely available software packages, and we detail our configuration to produce useful data for CBCT reconstruction testing.

2.1 The Role of Computed Tomography in Radiotherapy

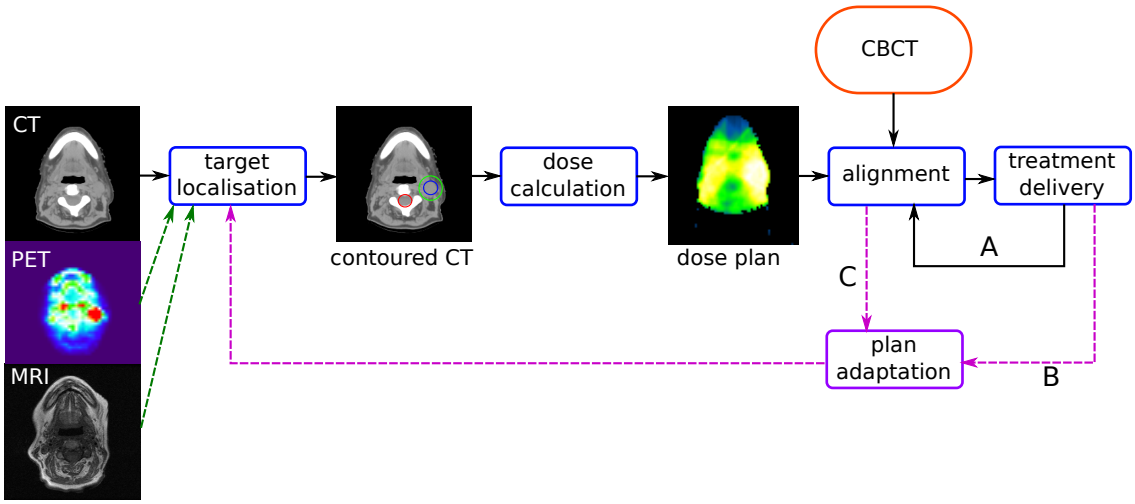


Figure 2.1: Radiotherapy flow diagram: the black arrows show a basic planning and treatment procedure; green arrows show additional modalities that may enhance localisation; and pink arrows show potential for adaptation from CBCT images. Images were obtained from the Cancer Image Archive [1, 2].

A flow diagram of a radiotherapy planning and treatment procedure is shown in Figure 2.1. At the beginning, a CT scan is acquired of the region of interest — the neck in the case illustrated. From this CT image, known as the planning CT, organs at risk (OAR) and the planning target volume (PTV) are identified. Due to its relatively low tissue contrast, this may be supplemented with other imaging modalities that offer additional structural or functional information, such as magnetic resonance imaging (MRI), ultrasound or positron emission tomography (PET) [15].

With the structures of interest identified within the image, the treatment plan is then made to satisfy a prescribed dose [16] [17], whilst sparing sensitive tissues, such as the spinal cord. CT is almost uniquely used for the basis of this planning, since it may be mapped into electron density, which determines the amount of interaction from the treatment beam.

At the time of treatment, which will likely be several days since the planning scans were taken, some degree of alignment is required to ensure the designed dose will be deposited into the correct region. It is at this point that CBCT is most commonly employed, where it can be used to register the planning CT to the patient's current position on the treatment couch. After satisfactory alignment, the treatment will then be delivered, and

will commonly be reapplied many times over the course of several weeks, all according to the initial plan [15] — this is illustrated with path A in Figure 2.1. This splitting of the delivery into many sessions is known as fractionation, and a typical treatment could be over 40 fractions or more [15].

The potential of the CBCT imaging information may influence the adaptation in several ways. Firstly, any change in anatomical structure between planning CT and CBCT, may be used to adapt the image localisation and hence dose calculation. This compensation may be made by some non-rigid registration of the planning CT, either between subsequent scans or at the time of treatment, shown respectively as paths B and C in Figure 2.1.

Another potential approach is to perform the adaptation and recalculation directly on the CBCT itself. The goal of this work is to unlock this potential, by sufficiently enhancing the quantitative accuracy of CBCT.

2.1.1 Basic Physics of Radiotherapy

Radiotherapy, as with X-rays used for the purposes of imaging, causes damage to tissues through ionisation. Ionisation is the creation of a charged atom or ion, due to the removal of its negatively charged electrons. When the atom is part of a molecule, this ionisation has the potential to alter the chemical bonds, and denature its structure. With DNA molecules, this can lead to the breaking of their strands, and ultimately cell death [24]. DNA can also be damaged through indirect mechanisms, whereby ionised water molecules denature it through interactions. When this occurs in a cancerous tumour, it can impede its reproduction or lead to its extinction.

There are several sources of ionising radiation that are used for cancer therapy. In Chapter 5, we will analyse three: high energy electromagnetic radiation (photons), hydrogen nuclei (protons), and carbon ions. Representative dose deposition characteristics for these three types are shown in Figure 2.2 through high density polyethylene (HDPE) — having a similar density to human soft tissues (0.98 g/cm^{-3}). To generate these curves, particle data was obtained from the NASA Space Radiation Laboratory [25], and photon data was simulated in the software package Gate [26].

Firstly, photon therapy is the most common form of treatment radiation, and is also generated by accelerating electrons into a heavy metal target. These electrons typically

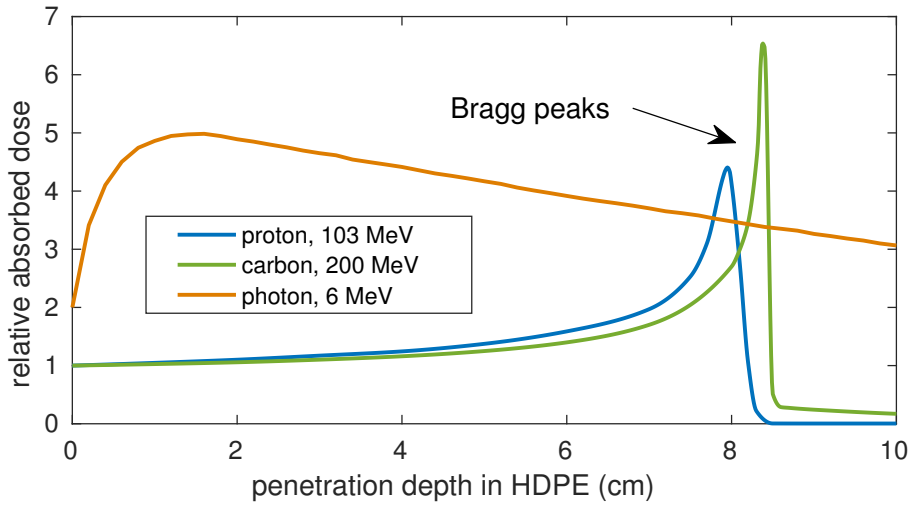


Figure 2.2: Illustration of energy deposition curves for different radiation types through HDPE.

have energy in the range 4–25 MeV, where the unit electron volt (eV) is equivalent to 1.602×10^{-19} Joules. To achieve these high energies, a linear accelerating wave-guide is employed before collision [27]. The profile shown in Figure 2.2 is generated from 6 MeV electrons, which gives rise to a polyenergetic spectrum after collision with a target. Although the absorption of these high energy photons is low, as they propagate through an object, they undergo Compton scatter events that will lower their energy and increase the probability of interactions. This is why the dose deposition initially rises in Figure 2.2, before slowly diminishing as the photons are attenuated. Apart from the initial rise, it can be seen that the dose deposition is reasonable constant throughout the material with photons.

Another class of radiation is ions, like the carbon and protons shown in Figure 2.2, which are accelerated directly into the specimen. From the dose deposition profiles, these clearly exhibit a significantly different interaction than from photons. They maintain a low entry dose for some distance into the object, before rising sharply in a peak, and falling rapidly to almost zero. This characteristic – known as the Bragg peak – arises since the interaction probability rises with a decrease in particle energy and this energy loss is inversely proportional to its velocity squared [28]. Therefore, the particles will rapidly come to rest and deposit most of their energy at some point as they are slowed along their trajectory. The mechanism of particle interaction with matter is primarily from Coulombic events, whereby electrons are ejected from atoms by the high energy particles, which themselves undergo small deflections. This is similar to Compton events described

in Section 2.2.2.

The location of the Bragg peak is dependent on both the atomic mass and the energy of the particle, which can be tuned to carefully deposit this energy into a target. Particle beams of varying energy can be superimposed to cover the whole tumour area [28]. The potential advantage of ions over photons, is that the dose can be much more localised, and healthy tissues, especially behind the tumour, will receive considerably less toxic radiation.

2.1.2 Computed Tomography for Treatment Plan Calculation

Unlike other medical imaging modalities, CT images have a unique role in, since they can be used as the basis for treatment planning calculation. This involves computing, through simulation, how the energy of the beam will be deposited in the specimen. This connection is intuitive, since the reconstructed image can be considered a 3D map of the attenuation strength of high energy electromagnetic radiation, and the basis for interactive calculation is exactly how the higher energy radiation will be absorbed by the specimen.

For photon therapy, the quantity to enable planning is the density of electrons, since they are the particles responsible for interaction with the treatment beam. For particle therapy, one instead uses the stopping power ratio (SPR) for the particle, which has a dependency on the atomic numbers of the material [29]. Although SPR is not equivalent to electron density, the two are very closely related [30]. From the CT reconstruction in terms of linear attenuation, the mapping into either electron density or SPR can be performed through a non-linear mapping. In practice, piecewise linear calibration curves are used for this process [30].

The calculation of the photon interaction itself is then done broadly by two techniques: convolution and Monte Carlo. In the case of Monte Carlo, individual photons are physically simulated through the specimen, whose interaction is probabilistically modelled from detailed interactive data, repeated for many photons, and the net energy deposition calculated as the superposition of all these events. The simulation is usually stopped at the point where the variance in regions of interest falls below some tolerance, but will typically require a huge amount of computational power. Popular examples of code for modelling physical interaction are PENELOPE [31] and Geant4 [32], which are both open source tools from the particle physics community. These codes are then exploited in

software giving interfaces for radiotherapy planning such as PRIMO [33] or GATE [26] respectively. A commercial example of Monte Carlo treatment calculation software are Elekta Monaco.

The other broad method of dose calculation relies on convolution, where appropriate point spread functions are applied to the primary beam within the specimen. Theses approaches are normally significantly faster than physical simulation, but suffer inaccuracies from spatially invariant assumptions of the convolution, though these can be reduced with heuristic perturbations. A particularly popular commercial convolutional method for photons is the Varian analytical anisotropic algorithm (AAA) [34], that will later be used for our real data dose experiment in Chapter 5. An open source package is then the matRad MATLAB toolbox [35], which allows the calculation for protons and carbon ions, and we will use this for the numerical dose testing in Chapter 5.

After calculation of the dose with a given beam set-up and intensity, the plan will then be checked to see if it satisfies constraints in terms of delivering a prescribed energy to the tumour, and limiting the dose to OAR [15–17]. If not, then adjustments or compromises are made before recalculation. This process can be generalised as a constrained optimisation procedure, and can be performed with mathematical techniques such as linear programming [36].

2.2 The Physics of Digital X-ray Imaging

X-rays are characterised as electromagnetic radiation with an energy ranging from 100 eV to 100 keV [37]. The formation of images from X-rays fundamentally depends on the physical interaction between electromagnetic radiation and matter, and its contrast can convey information about the object of interest. In order to probe these interactions, one requires a controlled method for the generation and detection of X-rays, which are also fundamentally dependent on interactions between radiation and matter.

The energy of an electromagnetic photon, which we will refer to throughout this section, is defined as

$$\xi = \frac{hc}{\lambda}, \tag{2.1}$$

where ξ is its energy in eV, $h = 4.136 \times 10^{-15}$ eVs is the Planck constant, $c = 3.00 \times 10^{8}$

m/s is the speed of light, and λ is the wavelength of the radiation in meters. Since hc is a constant, the energy is uniquely determined by its wavelength.

In this section, we will describe the life of X-rays as is relevant for commercial medical CT systems: through generation at the source, attenuation in the specimen, and finally the detection through scintillation and photo-absorption for measurement.

2.2.1 X-ray Generation

A simplified diagram of an X-ray source is shown in Figure 2.3.

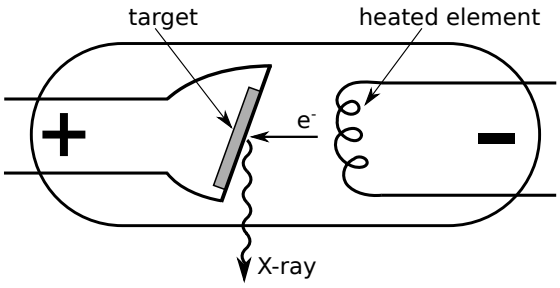


Figure 2.3: Diagram of X-ray tube with heated cathode element and tungsten anode target.

The assembly of the source is within a vacuum chamber, containing a heated cathode to produce electrons by thermionic emission, and an anode target. Due to the potential between the electrodes — rectified AC in the range of 70–140 kVp (indicating the peak voltage value) [37] — electrons are rapidly accelerated into the target material. To mitigate the large amount of heat generated at the anode, many commercial systems employ a rotating target [37], unlike the static system shown in Figure 2.3. With the exception of mammography, where molybdenum is used for the anode surface material [38], tungsten is the most common target material for X-ray imaging sources, since it has good X-ray generation efficiency and integrity at high temperatures.

Of relevance to CT imaging are the processes of interaction between the electrons upon impact against the target, giving rise to the energy spectrum of the source. The two processes resulting in the production of X-rays are bremsstrahlung and characteristic radiation. Firstly, bremsstrahlung taking its name from 'braking radiation' in German, is an interaction where the negatively charged electrons are deflected from their trajectory by the positively charged nuclei of the tungsten atoms [37]. These deflections usually

result in a reduction in kinetic energy, with the change in energy being emitted as a photon of radiation. With reference to Figure 2.4 showing a typical source spectrum from a potential of 120 kVp, the bremsstrahlung radiation is responsible for the continuous broadband hump profile. The intensity of this radiation is inversely proportional to the photon energy, up to a maximum value equal to the source potential. Since the lower energy photons are easily attenuated by the source window and filtration however, one sees a roll off in intensity for these energies, as is illustrated below 50 keV in Figure 2.4.

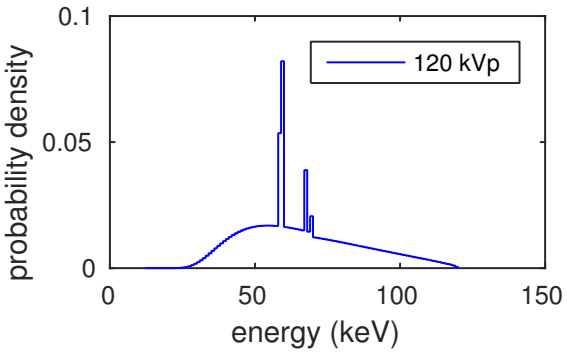


Figure 2.4: Spectrum of 120 kVp X-ray source from tungsten anode and 2mm aluminium filtration.

A second type of physical process leading to the generation of X-ray photons is characteristic radiation. This occurs when an electron collides with a low energy bound electron in the K-shell of a tungsten atom, causing it to be ejected. In this event, higher energy electrons readily transition to the vacant position, with the excess energy being released as a photon. As a quantum process, the energy of these photons can be only one of a few possibilities, with a value determined from the difference in energies between the quantum states. For example, the difference between L-shell and K-shell electrons in tungsten are 57.9 keV and 59.9 keV, giving rise to the first double peak in Figure 2.4; the other peaks are due to transitions from the M shell at 67.2 keV and 69 keV [37].

Along with the tube potential, which will determine the spectrum of the source, another factor is the tube current exposure. By adjusting the current in the heated element in Figure 2.3, the heating and therefore flux of electrons through thermionic emission is changed. To deliver a desired exposure, usually measured in mAs, one can both adjust this current, or operate the source in a pulsed manner.

2.2.2 X-ray Attenuation

Processes of attenuation that are relevant at the energies used in medical imaging are photoelectric and scattering events. In Figure 2.5, simplified illustrations of these processes are shown.

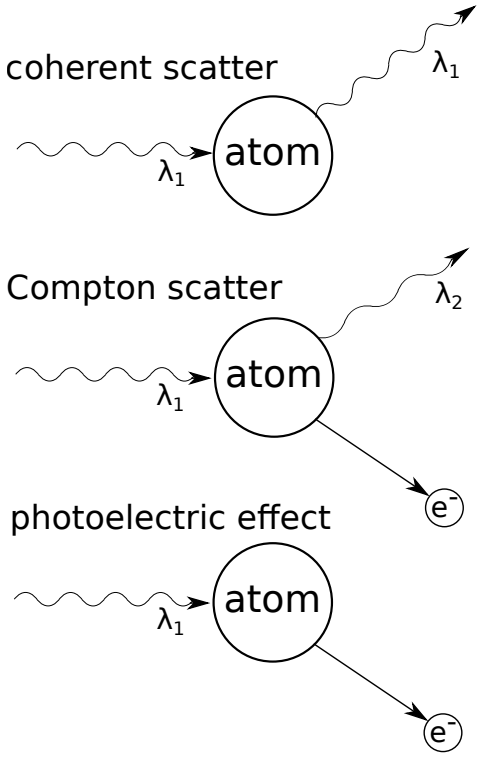


Figure 2.5: Simplified illustrations of X-ray attenuation mechanisms.

With reference to Figure 2.5, the various interactive mechanisms of X-ray radiation are defined as:

- Coherent scatter: analogous to elastic collisions in classical mechanics. The process involves an incident radiation causing excitation in a target atom, followed by the emission of radiation of the same wavelength λ_1, as the atom returns to a state of rest. Two variants of coherent scatter are Thompson and Rayleigh effects, where the excitation involves a single electron and all electrons in the target atom respectively [37].

- Compton scatter: also known as 'incoherent scattering' is typically a greater attenuation effect than coherent scatter [39]. It corresponds to an ejection of an electron from the atom, and the deflection of the incident photon. This scattered

radiation has a smaller energy (longer wavelength), since a portion is transferred to the kinetic energy of the electron. The angular distribution of Compton scatter is dependent on the incident photon energy and can be found from the Klein–Nishina formula [40].

- Photoelectric effect: is the absorption of the photon and ejection of one of the inner most (K shell) bound electrons in the target atom. Although not shown diagrammatically in Figure 2.5, this electron shell vacancy will subsequently be occupied by a higher energy electron, causing the release of characteristic radiation [37] as described in Section 2.2.1.
- Higher energy events: there are other mechanisms of electromagnetic radiation attenuation such as pair production, the annihilation of a photon and creation of an electron–positron pair, and photo-nuclear events, which involve the ejection of particles from a nucleus [37]. Since these only start to occur above energies of several MeV, they do not play a role for imaging X-rays, but can become active from the radiotherapy beam.

Material dependent X-ray attenuation

The mechanism that allows various regions in a heterogeneous specimen to be differentiated is their relative degree of X-ray attenuation. For biological tissues irradiated with a diagnostic X-ray source, the attenuating phenomena are the photoelectric and scattering effects [37], as illustrated in Figure 2.5. The combined attenuation strength of a given material can be quantified as [39]

$$\mu(\xi) = \rho N_g(\sigma_{\text{pe}}(\xi) + \sigma_{\text{incoh}}(\xi) + \sigma_{\text{coh}}(\xi)), \tag{2.2}$$

where ξ is the energy of the incident radiation, σ_{pe}, σ_{incoh} and σ_{coh} represent the interactive cross sections — quantifying the probability of interaction — of photoelectric, incoherent (Compton) and coherent (Rayleigh and Thompson) effects, ρ is the mass density, and N_g is the number of electrons per unit volume defined as

$$N_g = N_A \sum_i \frac{\omega_i Z_i}{A_i}, \tag{2.3}$$

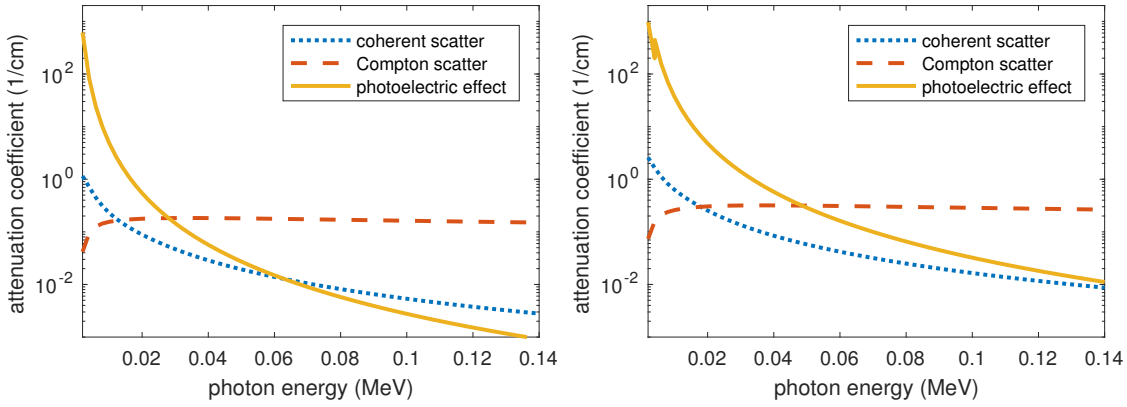
where N_A is Avagadro's constant (6.022×10^{23}), and Z_i, A_i, ω_i are atomic number, atomic weight and relative fraction by mass of a material's constituent elements [30].

A convenient parameter to use is the relative electron density, which is [30]

$$\rho_e = \frac{\rho N g}{\rho_{\text{water}} N_{g,\text{water}}}, \tag{2.4}$$

where $N_{g,\text{water}}$ is the absolute electron density of water and ρ_{water} its mass density, having a value of almost exactly $\rho_{\text{water}} = 1\,\text{g/cm}^3$ at room temperature and pressure.

Examples of the X-ray attenuation from the various physical processes for water and bone are shown in Figure 2.6. As can be seen, the attenuation profiles are both nonlinear with X-ray energy, and vary considerably between water and bone, which will commonly both be present in a medical scan.



(a) attenuation response for water

(b) attenuation response for bone

Figure 2.6: Attenuation responses for various processes for the cases of water and bone.

From (2.2), (2.3) and (2.4), one may model this nonlinear attenuation through the energy independent mass density or relative electron density — ρ and ρ_e respectively — given knowledge of the energy dependent interactive cross sections. This is possible by parameterising these processes, which is described in the next section.

A more common approach is to treat the attenuation as linear during the reconstruction process. This can be done either by ignoring the nonlinearity arising from the energy dependence of attenuation, or by compensating for it to some degree through calibration, which we will demonstrate in Section 2.6. This linear attenuation is often quantified

relative to water using Hounsfield units (HU) defined as

$$\text{HU} = 1000\frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}}. \tag{2.5}$$

The HU are not physically quantitative, as they will depend on the spectrum of the source and linearisation method applied. Nevertheless, conversion can be performed by generating a calibration table from scanning various known materials. Examples of calibration curves are shown in Figure 2.7. The calibration curves in Figure 2.7 are generated from 52 human

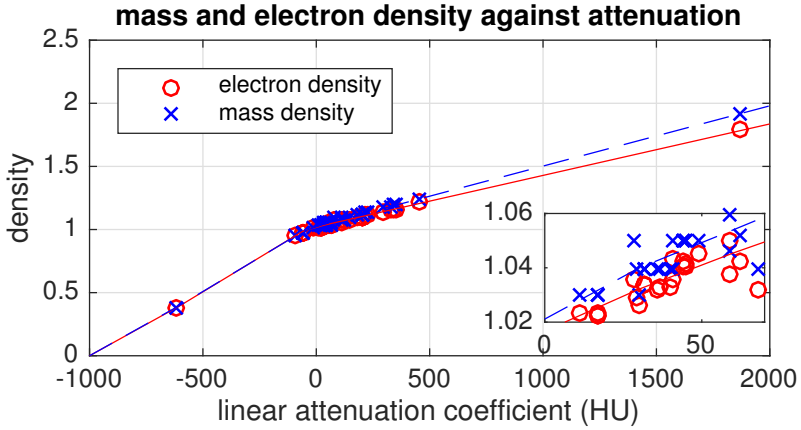


Figure 2.7: Plot of mass and electron density calibration curves against relative attenuation coefficient in Hounsfield Units (HU) at 60 keV, for range in biological tissues as defined in [3].

tissues defined in [3] and used in [41], according to a monoenergetic X-ray source at 60 keV. As in [30], the curves are piecewise linear, with 4 breakpoints at lung tissue, adipose tissue and around soft tissues, although some of these are difficult to discern in Figure 2.7. It can be observed that most tissues have very similar relative electron density and mass density, which deviates more significantly with higher attenuating materials, such as bone or metallic implants.

Existing Physical Parameterisations

Due to each cross-section in (2.2) being a non-linear function of energy and material, as shown in Figure 2.6, the total attenuation of a tissue is complicated and difficult to quantify exactly. However, this can be modelled given an accurate parameterisation of

(2.2). For example, Alvarez and Macovski [42] use the following model

$$\mu(\xi) = \underbrace{K_1 \rho_e Z_{\text{eff}}^{2.94} \xi^{-3}}_{\text{photoelectric}} + \underbrace{K_2 \rho_e f_{\text{KN}}(\xi)}_{\text{Compton scatter}}, \tag{2.6}$$

where $f_{\text{KN}}(\cdot)$ is the Klein–Nishina function [40] defined analytically in terms of energy as

$$f_{\text{KN}}(\xi) = \frac{1+\alpha}{\alpha^2}\left[\frac{2(1+\alpha)}{1+2\alpha} - \frac{1}{\alpha}\log(1+2\alpha)\right] + \frac{1}{\alpha}\log(1+2\alpha) - \frac{1+3\alpha}{(1+2\alpha)^2}, \tag{2.7}$$

where $\alpha = \xi/510.975$ keV. It describes the probability of Compton scatter for a single unbound electron in free space, though this will change for electrons in atoms [43]. K_1 and K_2 are scalar parameters to fit the model to data. It should be noted that the coherent scattering events are not explicitly modelled, though their contribution is low at the energies of interest [39].

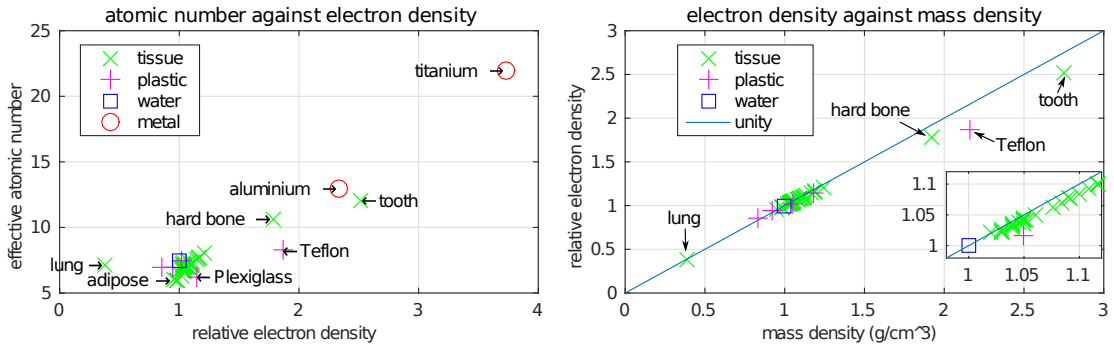
Z_{eff} is the effective atomic number for a composite material [44], which is calculated as

$$Z_{\text{eff}}^{2.94} = \sum_i \omega_i Z_i^{2.94}. \tag{2.8}$$

It can be interpreted as the equivalent average atom that a photon 'sees' during interaction with a mixture. For a single atomic species, (2.8) is the number of protons in its nucleus, or $Z_{\text{eff}} = Z$. There is not necessarily any relationship between ρ_e and Z_{eff}, but in general materials with a higher atomic number allow for a higher packing density of electrons.

According to (2.6), the energy dependent attenuation of any material may be parameterised by its effective atomic number and electron density. Several examples of these parameters for a range of differing material types are plotted in Figure 2.8a. Although there is clearly a strong correlation between the relative electron density ρ_e and effective atomic number Z_{eff}, in order to unambiguously model all classes of materials, at least two spectral measurements must be taken to separate the contribution from the two terms in (2.6) — since there is no one-to-one mapping from ρ_e to Z_{eff}. This is the basis of a 'dual-energy' CT technique, where from the model in (2.6) and measurements from two distinct X-ray spectra, the attenuation can be decomposed into photoelectric and Compton contributions, from which ρ_e could be unambiguously calculated [42].

In [11], the authors use a piecewise-linear fit from a monoenergetic equivalent attenuation to both of the energy independent factors in (2.6): $K_1 \rho_e Z_{\text{eff}}^{2.94}$ for photoelectric attenuation;



(a) (b)

Figure 2.8: (a) Plot of effective atomic number against relative electron density for a materials of: tissues from [3], plastics, water and metals; (b) relation between relative electron density and mass density for tissue, plastics and water, with unity line shown for illustration.

and $K_2\rho_e$ for Compton scatter. This fitting allows reconstruction from a single polyenergetic source. Due to the degradation of the energy dependent modelling in (2.6) at higher effective atomic number or higher energy however [39], this will not be quantitatively consistent in ρ_e between hard and soft materials; the model is also shown not to be consistent between synthetic and biological materials [11].

Another idea is to use physical materials as basis functions. For example, for biological specimens, water and bone may be considered [45]. The attenuation is then

$$\mu(\xi) = a_1\mu_{\text{water}}(\xi) + a_2\mu_{\text{bone}}(\xi) = \rho\left(a_1 m_{\text{water}}(\xi) + a_2 m_{\text{bone}}(\xi)\right), \tag{2.9}$$

where ρ is the mass density, $m_{\text{water}}(\xi)$ and $m_{\text{bone}}(\xi)$ are the energy dependent mass attenuation coefficients, and a_1 and a_2 can be binary [46] or water–bone fractions [47].

One may calculate ρ and map into ρ_e through a non-linear calibration if desired. The relation between mass and electron density is shown in Figure 2.8b. Although higher density materials deviate from that of water, as long as one can generate an estimate of the mass density, then the trend shown in Figure 2.8b can be approximated as piecewise-linear and conversion to electron density is possible, which is similar to the HU against density plot shown in Figure 2.7. A possible weakness of this model is the inaccuracies that will occur when tissues have a dissimilar mass-attenuation profile to both water or hard bone, such as adipose tissue [30].

2.2.3 X-ray Detection

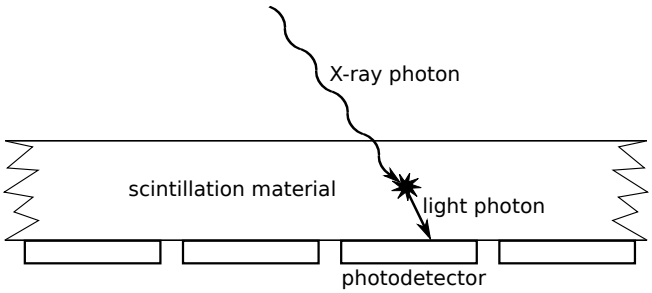


Figure 2.9: Simplified diagram of X-ray photodetector, with layer of scintillation material and array of detector elements.

After transmission through the specimen, the intensity of the X-rays must be digitally measured. To do this, a common approach is to position an array of photodetectors — usually either CCD or CMOS [48] [49] — behind a layer of scintillation material that re-emits absorbed X-ray photons as light. This light can then be absorbed by a photodetector and create a conduction electron, which can be measured through electronic amplification and analogue–to–digital converter (ADC).

For this type of detector, its response to various energies of X-ray photon can be modelled by considering the energy absorption of the scintillation material. To estimate this, we note the flux not absorbed by the scintillation is given as

$$I_{\text{out}} = I_{\text{in}} \exp(-d\mu(\xi)),$$

where d is the thickness of the scintillator, and $\mu(\xi)$ is its attenuation coefficient. With this, the quantity of photons having been absorbed by the material is

$$I_{\text{out}} - I_{\text{in}} = I_{\text{in}}\{\exp(-d\mu(\xi)) - 1\}.$$

The response function of the detector may now be derived as the photons absorbed by the scintillator, multiplied by their energy, and multiplied by some scalar constant relating the conversion from X-ray energy to a bit on the ADC. Finally, this can be expressed as [50]

$$\mathcal{R}(\xi) = K\xi \frac{\mu_{\text{en}}(\xi)}{\mu(\xi)}\{\exp(-d\mu(\xi)) - 1\}, \tag{2.10}$$

where $\mu_{\text{en}}(\xi)$ is the energy absorption coefficient of the scintillator.

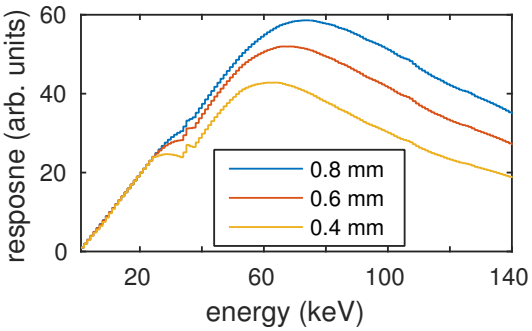


Figure 2.10: Scintillator response functions for CsI absorption of varying thickness. Response is relative to a simulated measurement after quantisation at the ADC, which has arbitrary physical units.

Examples of detector response functions calculated according to (2.10) for different thicknesses of scintillator is plotted in Figure 2.10.

The model for (2.10) is not exact for a true detector, since in reality there exists an angle dependence in CBCT, which will vary as one changes the angle of the ray. These effects may also be incorporated in an analytic model [51]. A more accurate model still, would include the mechanism of scintillation and photo-absorption and their spectral, angular and spatial dependencies.

2.3 Computed Tomography System Overview

The first commercially viable CT system was invented by Sir Godfrey Hounsfield in 1967, whilst working at EMI [52]. This first version consisted in a single source and detector, that were first translated laterally to form one projection, then rotated throughout 180° [53]. This system can be referred to as a single slice parallel beam CT. To reduce the scanning time, fan-beam systems were then developed, with an array of detector elements, which could acquire an entire projection without having to translate the source. The most common geometry used in medicine today is multi-slice fan-beam CT, where the detector is a 2D array of elements, allowing the acquisition of multiple axial slices in a single rotation of the gantry. With their enclosed gantry, multi-slice systems can rotate safely very quickly, typically ~ 0.5 seconds per cycle for popular commercial systems [54]. To cover larger regions along the patient's body, helical scans can be taken [55].

CBCT can therefore be considered an instance of multi-slice fan-beam CT. However, they are usually distinguished by having an exposed gantry, flat panel detector, and will

normally cover the entire field of view within a single acquisition. The first CBCT system for use in radiotherapy was developed by Jaffray and Siewerdsen in 2000 [56]. Due to the hazards from the heavy rotating components, the rotation times of clinical systems is usually ~ 1 minute, so considerably slower than enclosed gantry systems.

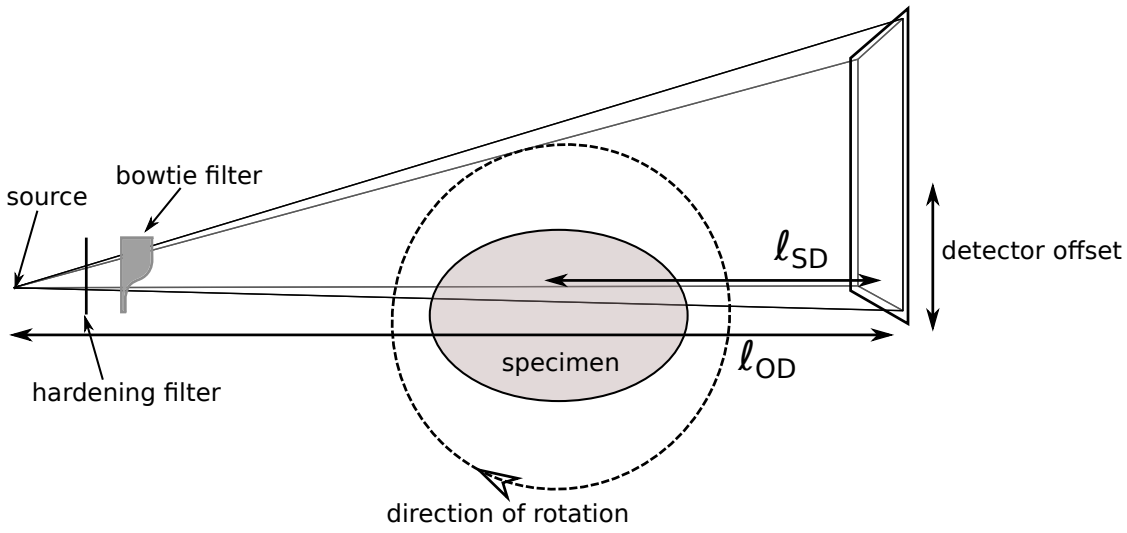
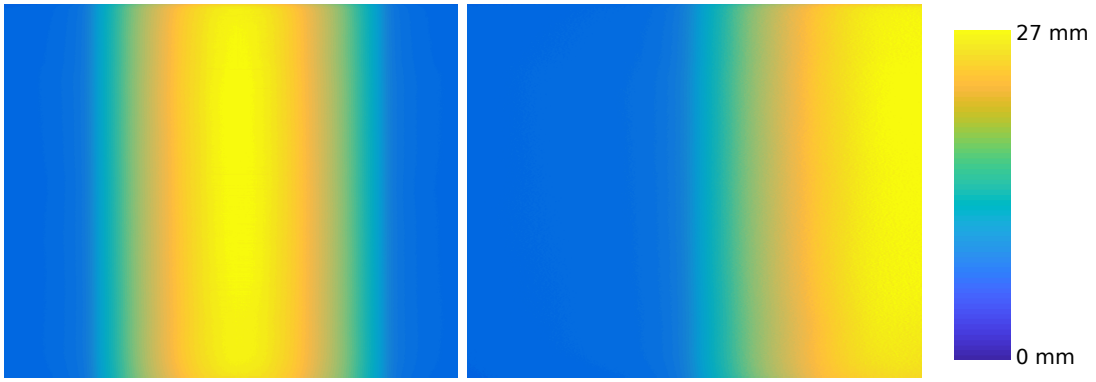


Figure 2.11: Profile view of CBCT geometry with offset detector, bowtie and beam hardening filters.

The geometry of a CBCT system is shown in Figure 2.11; a photograph of a clinical system is shown in Figure 1.1. In this case, the detector is a plane on the opposite side of the specimen to the source, which both rotate around the specimen with a fixed relative position. Due to the limited size of the detector, it is common practice to operate a scanner with an offset detector position, which effectively extends the size of the field of view, whilst allowing reasonable sampling throughout a 360° rotation [57].

Common elements of the beam path are hardware filters, such as the hardening and bowtie filters shown in Figure 2.11. Firstly, the hardening filter is a thin sheet of metal such as aluminium or titanium. Its effect is to remove many of the low energy photons from the spectrum of the source. This has the advantages of limiting the dose to the patient's skin, where the majority of these would otherwise be absorbed, and by improving the linearity of the X-ray attenuation [37]. Secondly, another very common element is known as a bowtie filter, which has the effect of shaping the source flux into the specimen. Examples of typical profiles from such filters are shown in Figure 2.12 — these are taken from the calibration images of a Varian TrueBeam On-Board Imager, and show the spatial variation in intensity as received at the detector plane.



(a) full fan profile (b) half fan profile (c) colour map

Figure 2.12: Source flux profiles from bowtie filtration for full fan (a) and half fan for offset detector (b).

Bowtie filters are very useful in CBCT. They have the effect of putting a significantly higher flux into the centre of the specimen, which achieves a more uniform SNR across a projection. This is valid for the typical structure of the anatomical sites treated in medical imaging, which for sections of the human body such as pelvis or chest, are well approximated as elliptical tubes. This has advantages of reducing the dose to the skin, and reducing the amount of scatter contaminating regions with the lowest SNR.

Another component that is used in some commercial CBCT systems is a scatter grid. These are positioned just in front of the detector panel, and consist of an array of thin lead strips orientated parallel to the source focal point. This has the effect of attenuating photons with different trajectories than the line-of-sight arising from scatter. A potential limitation of these is that in order to allow for offsetting the detector in a half fan mode, one can only collimate in one detector direction. On top of this, it has been shown that their use in CBCT may reduce tissue contrast, due to their inevitable attenuation of the signal of interest, which must be compensated through a higher radiation dose [58].

In X-ray CT, one is able to infer a specimen's attenuation through the drop in radiation intensity after transmission. The magnitude of this is found from the Beer–Lambert law, given for a monoenergetic beam as

$$I_{\text{out}} = I_{\text{in}} \exp\left(-\int_L \mu(\ell) d\ell\right), \tag{2.11}$$

where I_{in} is the incident intensity, L is the line-of-sight path of the beam through the specimen, ℓ is the vector pointing along L, I_{out} is the output intensity one is able to measure. Since in practice, μ is energy dependent and typically the source is polyenergetic,

the output intensity becomes

$$I_{\text{out}} = \int_E I_{\text{in}}(\xi) \exp\left(- \int_L \mu(\ell, \xi)\, d\ell \right) d\xi. \tag{2.12}$$

For a finite number of photons, the measured intensity will be probabilistic with a Poisson distribution, which may be expressed as

$$y_i \sim \text{Poisson}\left\{ \int_E b_i(\xi) \exp\left(- \int_{L_i} \mu(\ell, \xi)\, d\ell \right) d\xi + s_i \right\} \text{ for } i = 1, ..., N_{\text{ray}}, \tag{2.13}$$

where E is the energy spectrum, N_{ray} is the number of CT measurements, and $\boldsymbol{s} \in \mathbb{R}^{N_{\text{ray}}}$ is the expectation of the scatter or other background noise reaching the detector.

Although (2.13) would be accurate if the detectors measured individual X-rays hitting the detector, most commercial detectors as that described in Section 2.2.3, employ scintillators. Here, the detected optical photons take the form

$$y_{\text{det}} = A \sum_{n=1}^{y_i} z_n, \tag{2.14}$$

where y_{det} is the detected optical photons, A is the efficiency of the detector's ADC, and z_n are the number of optical photons generated from each of the y_i X-ray photons. Since both y_i and z_n are Poisson distributed, it follows that the y_{det} follows a compound distribution.

Instead of adopting the compound Poisson model, a common approximation is to instead use y_i in place of y_{det}. In this case, the effect of the detector response and resolution of the ADC can be encapsulated in the term $b(\xi) = \mathcal{R}(\xi) I_{\text{in}}(\xi)$.

Although we have shown the scatter term s as a scalar without statistical dependence, in practice it is highly dependent on the attenuation field μ, the source b and also the line of sight paths L. Another factor to consider, especially relevant in CBCT systems with a bowtie filter such as those shown in Figure 2.12, is the spectral and spatial variation of the source flux, though this may be represented with the scalar field $b_i(\xi)$.

The measurement and noise model in 2.13 is commonly used in modelling for reconstruction [11,46], and is what we will treat throughout this work. Although the noise is modelled solely as Poisson, in reality the electronics will contribute Johnson thermal noise — well approximated as additive white Gaussian — and the ADC will inevitably

result in quantisation noise [59]. Whilst this approximation may be reasonable [60], considering more accurate noise models could be an area of future extension.

2.4 Fundamentals of CBCT Reconstruction

In this section, we will cover some fundamentals of CBCT reconstruction. We will give details on the sampling from CBCT measurements, and filtered backprojection (FBP) technique, which is currently the dominant method employed in commercial systems [61]. We will use FBP as a basis of comparison in both Chapter 3 and Chapter 5.

Additionally, we will describe the discretisation of the specimen, which is a critical element for implementing any reconstruction on digital hardware. This will include the discretised version of the measurement model in (2.13), that we will use throughout all the experimental works in this thesis.

2.4.1 Cone-beam Sampling Insufficiency

The formation of CBCT projections can be considered as a mapping from the object to measurement space, and reconstruction is its inversion. Ideally, we would like this process to be both unambiguous and stable. For unambiguity, one requires that the measurements fully sample the entire object.

The sufficiency condition for reconstruction from cone-beam measurements was given by Tuy [18] and Smith [62], which refer to the sampling along source trajectories for a given object. This condition for the flat panel scan geometry of Figure 2.11 is:

Every plane through the object must intercept the source trajectory curve at least once.

There are many possible source trajectories for which this condition is satisfied, such as the helical scan used in multi-slice enclosed gantry CT, two adjacent circles, or a circle–plus–line. However, for the circular scan we treat in this thesis, this condition is violated, since many planes through a 3D object do not intersect the trajectory. This would imply that exact reconstruction of any 3D object is not possible.

A geometrical interpretation of this deficit was given by Grangeat [63], who presented an inversion formula for arbitrary source trajectories, by relating the projections to the so-

called Radon domain of the object. In this space, each projection onto a planar detector can be mapped to a hollow sphere, and the sampling from a continuous circular scan fills a torus. This is shown illustratively in Figure 2.13.

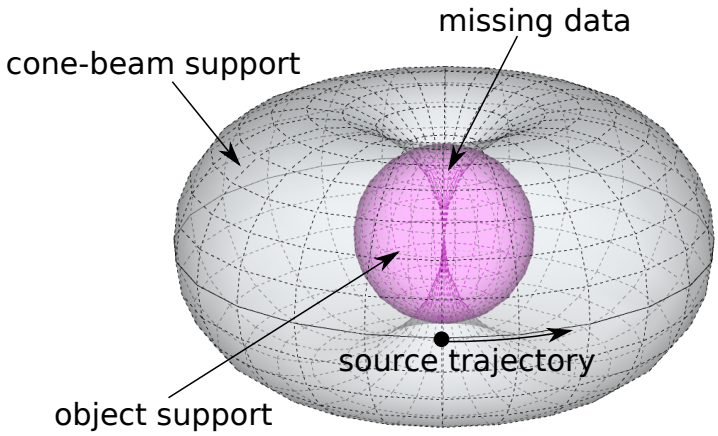


Figure 2.13: Illustration of sampling deficit in Radon space from a circular CBCT source trajectory.

In Figure 2.13, it can be seen that above and below the central slice, there is a progressively large hole in the sampled Radon space. How this deficit manifests itself in terms of the images was demonstrated in [64] as missing cones of data in the local Fourier transform of the object.

2.4.2 Filtered Backprojection

The filtered backprojection (FBP) technique is an efficient mapping between CT projections and an image. It is illustrated in 2D in Figure 2.14, and described in detail in [55, 65].

The FBP method is active on linearised projections of the object, which can be found from X-ray measurements in the monoenergetic case as $p_i = \log(I_{\text{in}}/I_{\text{out}})$. As the name suggests, these projections are firstly high-pass filtered, then backprojected along the trajectories of incident photons at each angle. This may be expressed for the 2D case shown in Figure 2.14 as

$$\mu(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} P(w, \theta)|w| \exp(j2\pi wt)\, dw\, d\theta, \tag{2.15}$$

where $P(w, \theta)$ is the Fourier transform of a projection. The inner integral applies the ramp filter in the Fourier space, whilst the outer integral performs the backprojection. For this

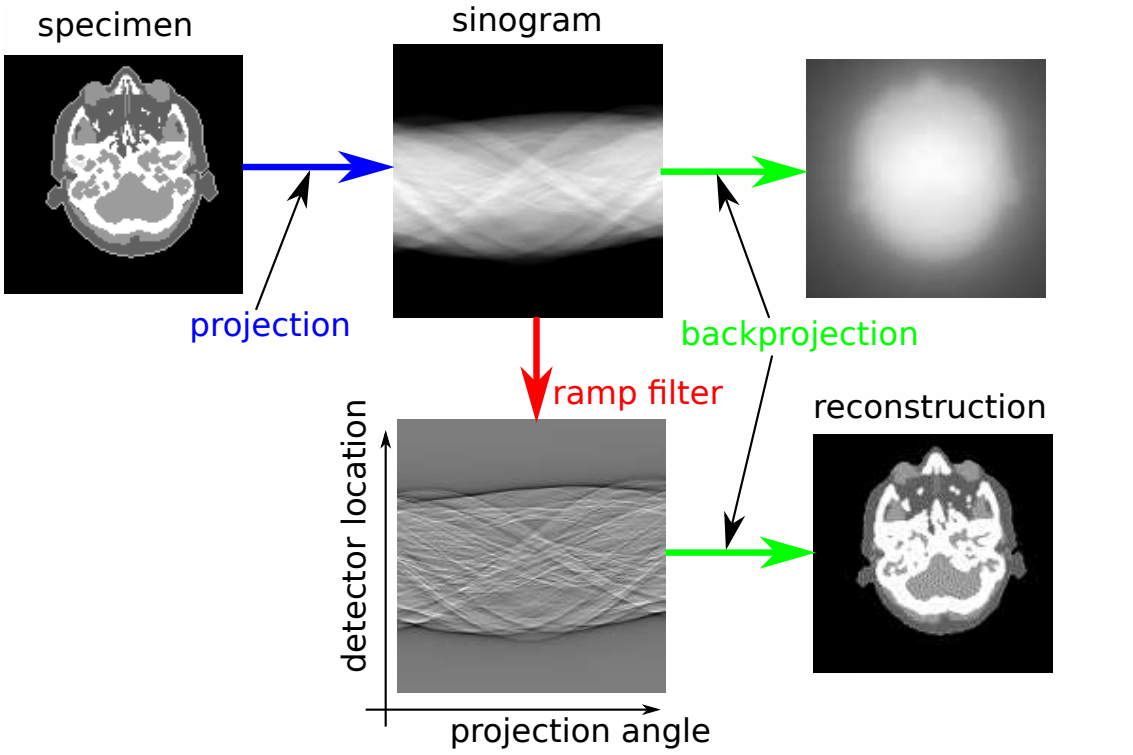


Figure 2.14: Diagrammatic representation of filtered back-projection process.

2D case, (2.15) is exact.

For 3D circular scanned CBCT, Feldkamp–Davis–Kress (FDK) [9] successfully applied FBP, with a modification to appropriately weight both the length and magnification of each ray. However, due to the fundamental insufficiency of this scan geometry, this approach is not exact, even for an infinite number of projections. It has been shown that FDK is exact for the interior of the torus in Radon space in Figure 2.13, and an extension has been made to also include data on its shell [66].

Despite the known deficit of circular scan CBCT and FDK, it still produces reasonable images in practice.

2.4.3 Discretisation Schemes

In order to use data from the CT system with a computer, discretisation is required. Since there are finite detector elements, and finite projection angles, the measurements are already discretised. On the other hand, one requires a representation of the underlying specimen in discrete space. For CT this is of critical importance as it dictates the form

of the projection operators, against which both the accuracy and computational cost are very sensitive [67].

Let us consider this discretisation of the continuous attenuation function μ in 3D as [55]

$$\mu(x, y, z) \approx \sum_{j=1}^{N_{\text{vox}}} \mu_j b_j(x, y, z), \tag{2.16}$$

where $b_j(x, y, z)$ represents the system basis functions associated with individual voxels, and μ_j is the discrete attenuation coefficient at the j^{th} voxel location, which is assumed constant over the area covered by the basis function. For cuboidal voxels, as is most common, this basis function has the form

$$b_j(x, y, z) = \text{rect}\left(\frac{x - \tilde{x}_j}{\Delta x}\right) \text{rect}\left(\frac{y - \tilde{y}_j}{\Delta y}\right) \text{rect}\left(\frac{z - \tilde{z}_j}{\Delta z}\right) \text{ for } j = 1, \ldots, N_{\text{vox}}, \tag{2.17}$$

where $(\tilde{x}_j, \tilde{y}_j, \tilde{z}_j)$ are the coordinates of the centre of the j^{th} voxel, and $\Delta x, \Delta y, \Delta z$ are the physical dimensions of a volumetric element. Another choice for basis function is to use radial blob functions, that exhibit a rotational invariance [68].

From the basis function discretisation, one finds a projection may be expressed as

$$p_i = \int_{L_i} \mu(x, y, z)\, d\ell = \sum_{j=1}^{N_{\text{vox}}} \phi_{i,j} \mu_j, \tag{2.18}$$

where we have that

$$\phi_{ij} = \int_{L_i} b_j(x, y, z)\, d\ell = [\mathbf{\Phi}]_{i,j}. \tag{2.19}$$

From this, we have a connection between the continuous space and the discrete domain, allowing us to rewrite the X-ray measurement model from 2.13 as

$$y_i \sim \text{Poisson}\left\{ \sum_{j=1}^{N_\xi} b_i(\xi_j) \exp\left(-[\mathbf{\Phi}\boldsymbol{\mu}(\xi_j)]_i\right) + s_i \right\} \text{ for } i = 1, ..., N_{\text{ray}}, \tag{2.20}$$

where N_ξ are the number of discrete energies used, of which both incident intensity and attenuation coefficient are a function.

A consequence of using cuboidal basis functions for discretisation is that in practice, calculating the entries of the matrix exactly as in (2.19) results in undesirable artefacts [69], which becomes particularly apparent in iterative reconstruction [67]. For this reason,

approaches that instead calculate the project matrices considering the finite detector size as in [70] offer significantly better performance.

In CBCT, both N_{ray} and N_{vox} can be of the order of 1×10^6. With this, $\boldsymbol{\Phi}$ is infeasibly large for storage in computer memory. Therefore, in practice $\boldsymbol{\Phi x}$ and $\boldsymbol{\Phi}^T \boldsymbol{p}$ are performed with functional operators, with their results calculated on the fly.

Another key consideration for numerically testing discretised systems, is to avoid committing the 'inverse crime' [71, 72]. This is when the operator used to synthesise the measurements $\boldsymbol{\Phi}$, is the same as that used during reconstruction. Since any physically underlying object will instead be continuous, this will ignore the approximation from the discrete conversion, and tested methods are likely to perform unreasonably well. Techniques to avoid this include either synthesising from a higher resolution specimen than that used during reconstruction, or using a highly detailed physical model, such as that we detail in Section 2.7.

2.5 Model Based Computed Tomography Reconstruction

From the fundamentals of reconstruction, we can now explore physical model based iterative reconstruction. By attempting to better model practical systems, they have the potential to operate a lower doses, produce more accurate images, or compensate for potential artefacts [59]. In recent years, most CT manufacturers have introduced methods of this type [73].

After providing a motivation for these methods, we will explore typical optimisation approaches that are used, and describe various regularisation techniques to enhance their performance for certain scenarios.

2.5.1 Motivation for Model Based Techniques and Low Dose Imaging

The techniques described in Section 2.4, such as the FBP and algebraic iterative methods, are solutions to an idealised X-ray system. Although we have briefly described the implications of additive noise, one could in principle overcome this by acquiring a sufficient number of very high flux measurements. Since an imaging X-ray is ionising radiation, as with the radiotherapy beam, there is strong motivation to operate at a lower dose.

The amount of radiation can be quantified as the effective dose in units of sievert (Sv), representing the average health risk from ionising sources. An exposure of 1 Sv presents a 5.5% of causing cancer [74]. On average, people receive 2.4 mSv from the atmosphere [75]. Alaei and Spezi [76] reviewed the effective dose from standard protocol CBCT acquisitions, which was found to be as high as 24 mSv for a single pelvis scan. With this, each acquisition is equivalent to 10 years of normal environmental radiation exposure, and if such a CBCT were taken at each of 40 fractions, then one could expect more than 1 in 20 patients to develop a secondary cancer from the imaging alone. It is therefore preferable to perform the scanning at a significantly reduced dose, if their use is to be increased.

As well as operating scanners at a low exposure, there may also be cases where collecting limited numbers of projections could be beneficial: to avoid collisions with the couch in 'non-coplanar' radiotherapy [77]; to enable the use of CBCT during image guided surgery [78]; and when the projections are temporally resolved for dynamic imaging [79]. Additionally, fixed gantry systems offer the potential to greatly improve the temporal resolution, but often result in a reduced number of projections than rotated systems [80].

Although a low dose is desirable, the FBP and algebraic reconstruction techniques will suffer in this case [81], since they rely on idealised assumptions that no longer hold. Instead of trying to exactly invert a measurement system, a more robust approach is to seek a solution that is optimal, given a detailed physical model for the noise in the system, and incorporating appropriate prior information one may have.

2.5.2 Iterative Reconstruction for Noisy Measurements

Instead of trying to exactly invert a linear system of equations, one could minimise a cost function incorporating an appropriate noise model. A general form of this may be expressed as

$$\underset{\boldsymbol{\mu}}{\operatorname{argmin}}\, \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{y}), \tag{2.21}$$

where $\mathcal{D}(\cdot, \cdot)$ is some objective function evaluating the mismatch between image $\boldsymbol{\mu}$ and measurements $\boldsymbol{y}$. A straightforward way to minimise (2.21) is through gradient descent as

$$\boldsymbol{\mu}^{k+1} := \boldsymbol{\mu}^k - \lambda \boldsymbol{\nabla} \mathcal{D}(\boldsymbol{\mu}^k, \boldsymbol{y}) \tag{2.22}$$

The simplest approach of this type is to model the noise as additive white Gaussian, in which case an appropriate cost function is the least squares as

$$\underset{\boldsymbol{\mu}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{\Phi}\boldsymbol{\mu} - \boldsymbol{p}\|^2. \tag{2.23}$$

The objective function in (2.23) may then be minimised through many well established optimisation algorithms such as gradient descent, conjugate gradient descent, coordinate descent, or quasi-newton methods such as limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [82], all of which have been applied for CT reconstruction [83] [84]. The gradient descent optimisation of (2.23) is expressed as

$$\boldsymbol{\mu}^{k+1} := \boldsymbol{\mu}^k - \lambda\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\mu}^k - \boldsymbol{p}). \tag{2.24}$$

In the CT field this algorithm is commonly known as the simultaneous iterative reconstruction technique (SIRT) [65], which has a reputation for being very slow to converge. For the algorithm in (2.24), one requires a step size of $\lambda < 2/\sigma_{\max}(\boldsymbol{\Phi})$ to guarantee convergence, where $\sigma_{\max}(\cdot)$ is the maximum singular value of the system matrix. For typical CT matrices, this step size becomes extremely small, though preconditioning may be employed to improve upon this [85].

Since the noise is better modelled as Poisson as in (2.20), the Gaussian model on the log of the measurements implied in (2.23), will degrade performance for noise limited scans [60]. One can instead consider the underlying noise model to maximise a likelihood function with respect to $\boldsymbol{\mu}$. From the discretised independent Poisson measurements model in (2.20), the log-likelihood function can be expressed as

$$L(\boldsymbol{\mu};\boldsymbol{y}) = \sum_{i=1}^{N_{\text{ray}}} y_i \log\left[\sum_{j=1}^{N_\xi} b_i(\xi_j)\exp\left(-[\boldsymbol{\Phi}\boldsymbol{\mu}(\xi_j)]_i\right) + s_i\right] - \sum_{j=1}^{N_\xi} b_i(\xi_j)\exp\left(-[\boldsymbol{\Phi}\boldsymbol{\mu}(\xi_j)]_i\right) - s_i. \tag{2.25}$$

Making a monoenergetic approximation, the negative log-likelihood (NLL) can be expressed as

$$-L(\boldsymbol{\mu};\boldsymbol{y}) = \sum_{i=1}^{N_{\text{ray}}} b_i\exp\left(-[\boldsymbol{\Phi}\boldsymbol{\mu}]_i\right) + s_i - y_i\log\left[b_i\exp\left(-[\boldsymbol{\Phi}\boldsymbol{\mu}]_i\right) + s_i\right]. \tag{2.26}$$

This function is now continuous and differentiable with derivative

$$\boldsymbol{\nabla} L(\boldsymbol{\mu}, \boldsymbol{y}) = \boldsymbol{\Phi}^T \left\{ \boldsymbol{b} \odot \exp\left(-\boldsymbol{\Phi}\boldsymbol{\mu}\right) \odot \left[\boldsymbol{y} \oslash \left(\boldsymbol{b} \odot \exp\left(-\boldsymbol{\Phi}\boldsymbol{\mu}\right) + \boldsymbol{s} \right) - \mathbf{1} \right] \right\}, \tag{2.27}$$

which can be used as the basis of a gradient descent based optimisation, to maximise the likelihood of the underlying Poisson model. One feature of (2.26) however, is that the function is non-convex for non zero $\boldsymbol{s}$ [86]. In [86], the authors present a monotonically decreasing algorithm for minimising (2.26) despite this non-convexity. Although we have shown here the monoenergetic NLL, several extensions using the polyenergetic parameterisations from Section 2.2.2, such as the water–bone or photoelectric–Compton approximations have been made [11, 46, 47]. We will evaluate these methods in Chapter 3.

In an attempt to preserve convexity, an alternative approach is to opt for a weighted least squares objective function, which is a second order Taylor expansion approximation of the NLL in (2.26) [46]. This gives an objective function of the form

$$\underset{\mu}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{\Phi}\boldsymbol{\mu} - \boldsymbol{p}\|^2_{\boldsymbol{W}} \equiv \frac{1}{2} (\boldsymbol{\Phi}\boldsymbol{\mu} - \boldsymbol{p})^T \boldsymbol{W} (\boldsymbol{\Phi}\boldsymbol{\mu} - \boldsymbol{p}), \tag{2.28}$$

where $\boldsymbol{W}$ is a diagonal weighting matrix having entries

$$W_{i,i} = \frac{(y_i - s_i)^2}{y_i}. \tag{2.29}$$

This has also been extended beyond the idealised Poisson model to approximate the addition of electronic noise [60], and the effect of beam hardening correction [87]. The derivative of the objective in (2.28) is then

$$\boldsymbol{\nabla} \mathcal{D}(\boldsymbol{\mu}^k, \boldsymbol{y}) = \boldsymbol{\Phi}^T \boldsymbol{W} (\boldsymbol{\Phi}\boldsymbol{\mu}^k - \boldsymbol{p}). \tag{2.30}$$

Regularisation

Thus far, the optimisation based reconstruction methods presented, such as the gradient descent on the LS, WLS, or NLL objective functions, attempt to find a solution that is the best fit to the measurements. Taking a Bayesian interpretation, these can be considered as maximum likelihood (ML) approaches. To achieve better statistical performance, and avoid over-fitting, one often wishes to incorporate a prior distribution of the parameter to be estimated. For example, one could consider a maximum a posteriori (MAP)

estimation [88], which is achieved in practice by regularising the ML. A general form of this modifies the objective function to

$$\underset{\mu}{\operatorname{argmin}}\, \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{y}) + \mathcal{R}(\boldsymbol{\mu}), \tag{2.31}$$

where $\mathcal{R}$ is some regularisation function that is active on the image. If the regularisation function is differentiable, then reconstruction can still be performed through gradient descent, which has the form

$$\boldsymbol{\mu}^{k+1} := \boldsymbol{\mu}^k - \lambda \left(\boldsymbol{\nabla} \mathcal{D}(\boldsymbol{\mu}^k, \boldsymbol{y}) + \boldsymbol{\nabla} \mathcal{R}(\boldsymbol{\mu}) \right). \tag{2.32}$$

Classical choices for the regularisation is some penalty function between neighbouring pixels to promote smoothness, whilst attempting to preserve edges. A generic formula for this is

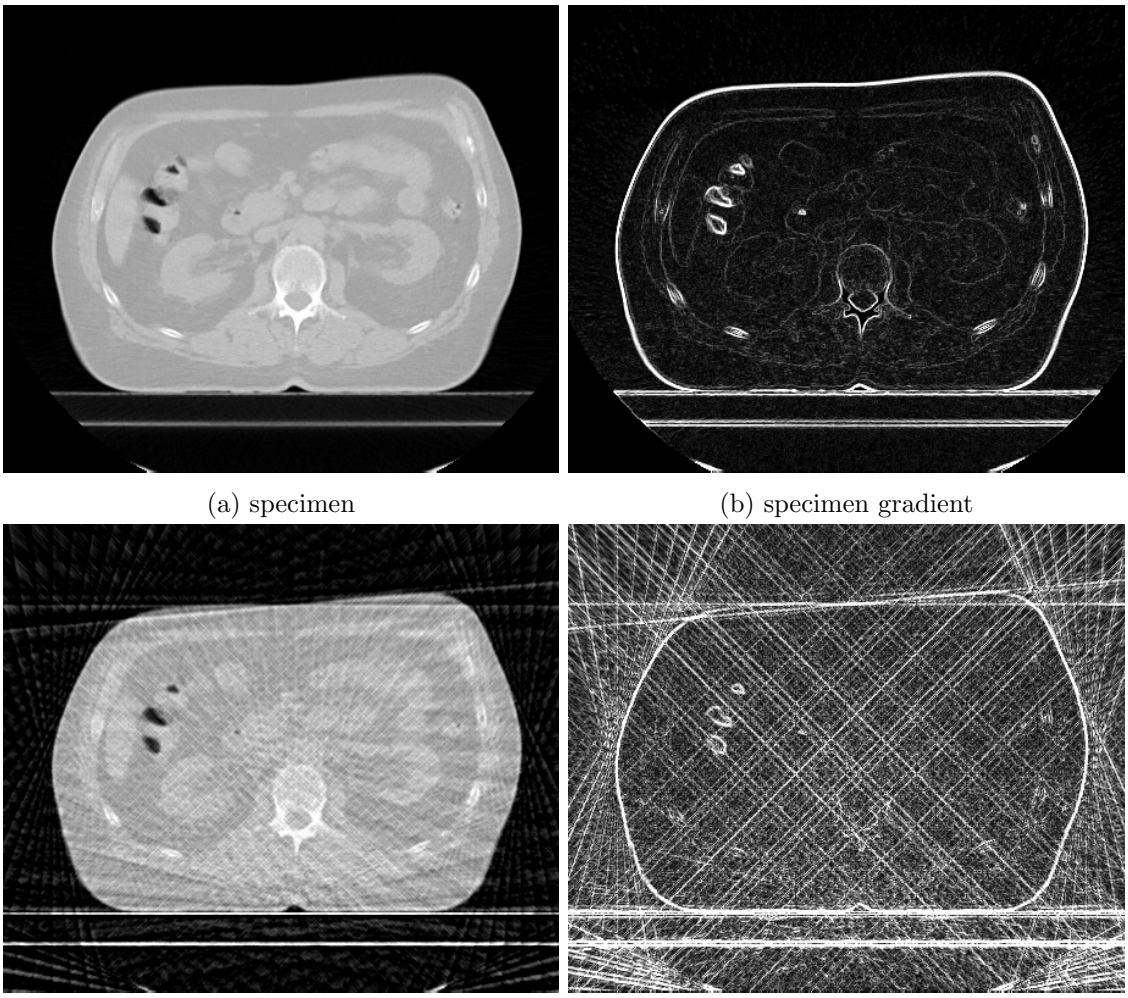
$$\mathcal{R}(\boldsymbol{\mu}) = \sum_{j=1}^{N_{\text{vox}}} \sum_{k \in \mathcal{N}_j} \Psi(\mu_j - \mu_k), \tag{2.33}$$

where choices for Ψ could be a simple quadratic, the pseudo-Huber loss [89, 90] and hyperbolic function [91], and $\mathcal{N}_j$ represents a neighbourhood of pixels.

2.5.3 Sparsity Based Regularisation

After the seminal work by Candes et al. and Donoho [92, 93] in 2006, a great deal of interest grew around the field of compressed sensing. Here, it was shown that from incoherent measurements, one can guarantee (with very high probability) to exactly reconstruct a sparse signal from very few samples. This was of great interest to CT, since it offered the potential to perform CT reconstruction from few projections. Additionally, it highlighted the benefit of non-differentiable regularisation to promote sparsity. Although CT measurement matrices do not strictly follow the required structure for compressed sensing, they have been demonstrated to have similar properties in some cases [94]. Nevertheless, the same algorithms have been employed with success in CT.

To motivate the use of sparsity for reconstruction, Figure 2.15a shows a high quality image from a CT scan of a male abdomen region. Although this image contains a good deal of structure, there appear to be only a few distinct grey levels, and there are significantly large regions of homogeneous intensity, corresponding to organs. This piecewise homogeneity



(a) specimen (b) specimen gradient

(c) FBP reconstruction (d) FBP gradient

Figure 2.15: Gradient sparsity from high quality CT image and crude FBP reconstruction. The data was obtained from a lung segmentation challenge [4] on the Cancer Image Archive [5].

means that the spatial variation, shown as a discrete gradient in Figure 2.15b, has a much lower proportion of significant values. On the other hand, from a poor quality parallel FBP reconstruction with only 36 projections shown in Figure 2.15c, the gradient in Figure 2.15d becomes highly structured. Therefore, if both Figure 2.15a and Figure 2.15b provided similar fits to the measured data, then selecting the candidate with the highest degree of sparsity in its spatial gradient would result in a far more desirable image.

A general form of sparsity based regularisation applicable to CT reconstruction addresses the unconstrained optimisation problem such as

$$\underset{\mu}{\operatorname{argmin}}\, \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{y}) + \alpha \|\mathcal{U}(\boldsymbol{\mu})\|_1, \tag{2.34}$$

where $\|\cdot\|_1$ represents the L^1-norm equivalent to the sum of all elements; $\mathcal{D}(\cdot,\cdot)$ is a differentiable data fidelity term such as the WLS in (2.28); $\mathcal{U}(\cdot)$ is some operator to transform the image into some domain in which it is sparse; and α is a regularisation constant to appropriately weight each of the terms.

Typical sparsity promoting transforms that have gained popularity in imaging are the image gradient, discrete wavelet transformations such as Haar [95] or Daubechies [96], and a learnt 'dictionary' [97]. Of these, the L^1-norm of the image gradient, known as total variation (TV) [98] has proved particularly popular, where one isotropic definition in 2D is given as

$$\|\mathcal{U}(\boldsymbol{\mu})\|_1 = \mathrm{TV}(\boldsymbol{\mu}) \equiv \sum_{x,y} \sqrt{(\mu(x,y) - \mu(x-1,y))^2 + (\mu(x,y) - \mu(x,y-1))^2}. \tag{2.35}$$

The reasoning behind this choice is that for homogeneous slabs of tissue with constant attenuation, the image is well approximated as being piecewise constant, which will have a high level of sparsity in the gradient space — as was illustrated in Figure 2.15. It has been shown that the level of gradient sparsity is linearly related to the number of required measurements for TV reconstruction [99]. Unlike classical regularisation, TV as in (2.35) is non-differentiable, so it cannot be solved with the gradient descent as given in (2.32).

One work that capitalised from the emergence of compressed sensing was that by Sidky et al. [100], where they attempted to regularise CT reconstruction with TV. The authors in this case used a differentiable approximation to the TV as in [101], making it admissible to the classical algorithm in (2.32).

Proximal Gradient Descent

More recently, general techniques have been developed to account for convex but non-differentiable regularisation functions, such as TV. One such method is proximal gradient descent, which may be expressed as

$$\boldsymbol{\mu}^{k+1} := \mathbf{prox}_{\lambda^{-1}\mathcal{R}}\left(\boldsymbol{\mu}^k - \lambda \boldsymbol{\nabla}\mathcal{D}(\boldsymbol{\mu}^k, \boldsymbol{y})\right), \tag{2.36}$$

where $\mathbf{prox}_{\lambda^{-1}\mathcal{R}}(\cdot)$ is the proximal operator is defined as

$$\mathbf{prox}_{\lambda^{-1}\mathcal{R}}(\boldsymbol{z}) = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} \frac{\lambda}{2}\|\boldsymbol{z}-\boldsymbol{\mu}\|^2 + \mathcal{R}(\boldsymbol{\mu}). \tag{2.37}$$

The effect of the proximal operator is therefore returning a point that minimises the regularisation function, whilst not moving too far from the starting location [102]. Further insight can be made through the fact that a solution $\boldsymbol{z}^*$ to (2.37) may be written as

$$(\boldsymbol{z}^* - \boldsymbol{\mu}) + \lambda\dot{\boldsymbol{\mathcal{R}}}(\boldsymbol{z}^*) = 0, \tag{2.38}$$

where $\dot{\boldsymbol{\mathcal{R}}}(\cdot)$ is the sub-gradient of the potentially non-differentiable regularisation function. Through rearrangement, this can be viewed as a sub-gradient descent, though with respect to the final value $\boldsymbol{z}^*$, and hence considered a 'backward step'. In combination 2.36 can be viewed as a gradient descent followed by a backward sub-gradient descent, so is commonly referred to as 'forward–backward splitting' [102].

A key feature of proximal gradient descent is that convergence is guaranteed for step-sizes of

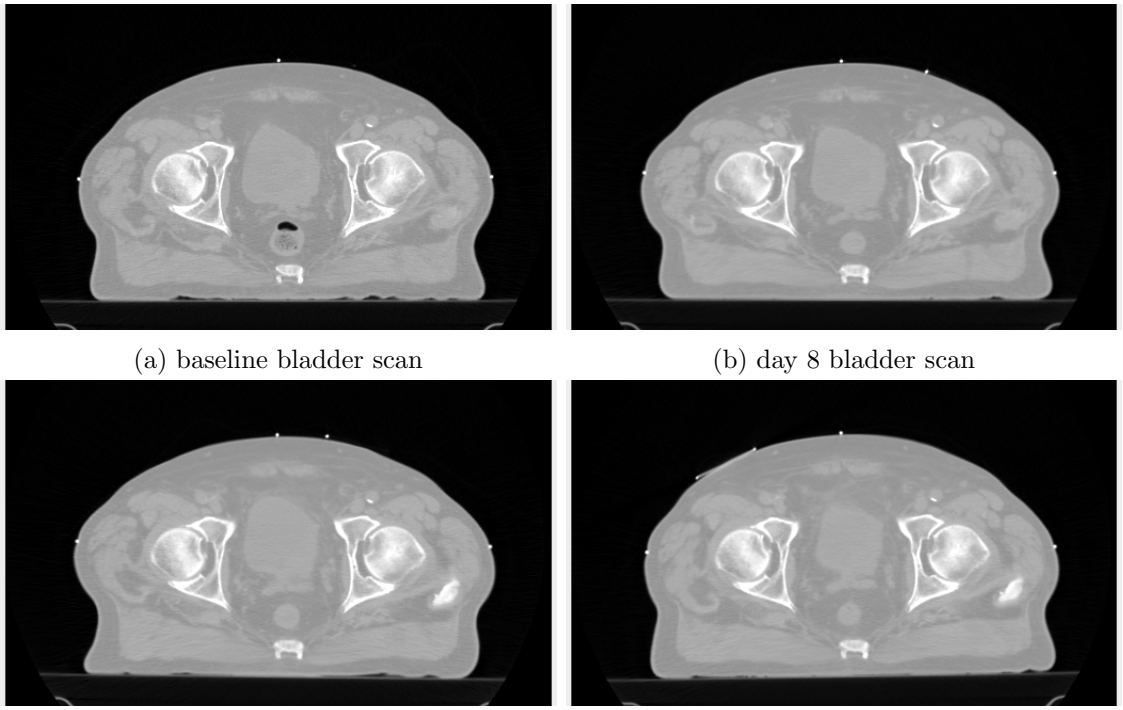
$$\gamma < \frac{2}{L(\boldsymbol{\nabla}\mathcal{D})}, \tag{2.39}$$

where $L(\boldsymbol{\nabla}\mathcal{D})$ is the Lipschitz constant of the gradient. Therefore, convergence is independent of the choice of regularisation function.

A popular approach for solving (2.37) in practice is that by Chambolle [103], which was accelerated considerably in [104], and applied to CBCT reconstruction along with a generalisation to other regularisation in [105].

Prior Regularisation

In the context of CT for radiotherapy, where any patient will receive a number of scans, one can exploit information between these to either reduce the amount of new measurements required or enhance the quality of the follow up scans. To illustrate this concept, Figure 2.16 shows 4 CT scans of a bladder patient over a 16 day period. The similarities between these successive images is clear, with the bones have practically identical structure in each image. The differences between the scans are slight deviations in body position, and critically for this application, the size and shape of the bladder and rectum at the



(a) baseline bladder scan (b) day 8 bladder scan

(c) day 13 bladder scan (d) day 16 bladder scan

Figure 2.16: Four successive images from a series of CT scans of a bladder patient [6]. The three subsequent images are rigidly registered to the bones of the baseline scan.

centre of the image. Since the differences between these successive scans is small, when compared to the total information in any single acquisition, one could exploit this sparsity against a prior image to regularise a model based reconstruction.

A general form of sparsity promoting prior regularisation is the solution of

$$\underset{\mu}{\operatorname{argmin}}\, \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{y}) + \alpha \|\mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{prior}})\|_1, \tag{2.40}$$

where $\alpha|\mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{prior}})|_1$ is some sparsity promoting regularisation function on a single or series of prior images $\boldsymbol{\mu}_{\text{prior}}$. One example of this was introduced as the prior image constrained compressed sensing (PICCS) [12], combining regularisation on the difference to another image and spatially within the same image. This allows one to still exploit the structure, such as sparse gradient in the image itself, whilst also incorporating redundancy from previous scans. A general form of PICCS is

$$\|\mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{prior}})\|_1 = \beta \|\mathcal{U}_1(\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{prior}})\|_1 + (1-\beta)\|\mathcal{U}_2(\boldsymbol{\mu})\|_1, \tag{2.41}$$

where β is a scalar in the range $[0, 1]$ to set the relative weighting of the two regularisation

functions. An example of this is when both regularisers are TV functions, where the expression becomes

$$\|\mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{prior}})\|_1 = \beta \text{TV}(\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{prior}}) + (1 - \beta)\text{TV}(\boldsymbol{\mu}), \tag{2.42}$$

One could instead consider combining the two regularisation functions into a single TV term that treats the prior image or images as an extra dimension, given in 2D as

$$\begin{aligned}&\text{TV}_p(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{prior}}) = \\ &\sum_{x,y} \sqrt{(\mu(x,y) - \mu(x-1,y))^2 + (\mu(x,y) - \mu(x,y-1))^2 + \beta(\mu(x,y) - \mu_{\text{prior}}(x,y))^2}.\end{aligned} \tag{2.43}$$

This is now equivalent to the spatial–temporal regularisation in [106].

The justification for employing some regularisation such as TV against a prior image is that if there is some small change between scans, such as the growth of a tumour, then this will be a piecewise constant difference having a sparse gradient. One critical factor however is the alignment of the images. Unless the consistent structures are spatially matched, such as the bones in Figure 2.16, then the difference between the images will also include redundant information. In order to overcome this problem, authors in [107] proposed the PIRPLE framework, where the registration of the prior image is updated as the reconstruction iterations progress. This was then extended in [108] from rigid to non-rigid deformations of the prior.

Another interpretation of the prior regularisation is that one forgoes the redundancy between scans by only finding the difference image. One could be explicit with this by writing the objective function as

$$\underset{\mu}{\text{argmin}}\, \mathcal{D}(\boldsymbol{d}, \boldsymbol{y}) + \alpha \|\mathcal{U}(\boldsymbol{d} - \boldsymbol{\mu}_{\text{prior}}, \boldsymbol{d})\|_1, \tag{2.44}$$

where $\boldsymbol{d}$ is the difference image simply defined as $\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{prior}}$. This concept was denoted as the reconstruction of difference (RoD) in [109]. Although it is equivalent to the original PICCS for a full scan, the concept allows one to reconstruct only a small region of interest where any differences are expected to occur, offering a considerable computational speed-up.

2.5.4 Accelerated Iterative Reconstruction

Although employing iterative reconstruction offers the benefit of better noise modelling and regularisation, it usually comes at a cost, which is a huge computational expense. A reason for this is clear from the PWLS algorithm in (2.32). As opposed to FDK with a fast filtering followed by a single backprojection, a single iteration requires both a forward and backprojection. Since CT matrices are typically very poorly conditioned, the allowable step-size for convergence becomes very small, and many iterations are required for convergence. With this, the computation of an iterative method can easily become several orders of magnitude greater than that of FDK [110]. There are however several approaches to lessen this gap.

Preconditioning

A natural approach to increasing the convergence rate is by improving the conditioning of the objective function through preconditioning. The most straightforward instance of such is with diagonal preconditioning, where each gradient term is scaled prior to backprojection. A popular instance of diagonal preconditioning is known as separable paraboloid surrogates (SPS) — also known as quadratic majorisation — which was introduced in [86] to provide a monotonic decrease of the NLL function for CT. This was later generalised and theoretically analysed in [111] for nonsmooth regularisation functions such as TV. In [46], SPS was also shown to be applicable for both PWLS and a polyenergetic statistical reconstruction method. This approach only offers modest acceleration however, as it does not fully exploit the structure of system matrix $\boldsymbol{\Phi}$.

Another approach that in some cases offers more dramatic acceleration is spectral preconditioning, as analysed for tomographic reconstruction in [85]. For the least-squares objective function in (2.23) with associated gradient step in (2.24), this attempts to compensate for the blur introduced by the application of $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$. This is achieved through

$$\boldsymbol{\mu}^{k+1} := \boldsymbol{\mu}^k - \lambda\mathcal{F}^{-1}\left(\mathcal{F}(\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\mu}^k - \boldsymbol{p})) \oslash \left[\mathcal{F}(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\delta}) + \nu\right]\right), \tag{2.45}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are 2D FFT and IFFT operations respectively, $\boldsymbol{\delta}$ is an 'impulse' that can be generated in practice by setting only the central most pixel to 1, and ν is a factor that should be increased in noisy scenarios for stability. Therefore, $\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\delta}$ represents

the point spread function (PSF) for the system operator. One potential limitation of this approach is that it assumes spatial invariance. This would restrict the application of weighting terms of PWLS for example, although Ramani and Fessler [112] showed through splitting the noise model from the system operator, that spectral preconditioning can be applied for CT. In CBCT, there is also high spatial variability throughout the image slices, where the PSF will only cover the central slice. Therefore, it is unlikely spectral preconditioning could be directly used [112].

Subset Gradient Descent

Another approach for acceleration is through calculating the gradient only with a subset of the data. This potentially leads to a natural speed up by a multiple of the down sampling factor, but will introduce some level of error, since one is approximating the true gradient. In machine learning, stochastic gradient descent methods are very commonly employed for their speed, where the subset is randomly selected from an often huge training set [113]. In CT, one often opts for deterministic selection schemes however, since they may lend themselves to more efficient computational implementation, and can even be superior to stochastic selection [7, 23]. This deterministic scheme is known as 'ordered subsets', and has been applied to the SPS for CT NLL in [114], to greatly enhanced its empirical performance.

One can write a prototypical subset gradient descent method as

$$\boldsymbol{\mu}^{k+1} := \boldsymbol{\mu}^k - \lambda \boldsymbol{\nabla} \mathcal{D}(\boldsymbol{\mu}^k, \boldsymbol{y}, \mathcal{S}^k), \tag{2.46}$$

where $\mathcal{S}^k \subset [1, \ldots, N_{\text{ray}}]$ is the set of measurements used in each subset. Taking equally spaced projections is denoted

$$\boldsymbol{\iota}^k = [l^k : N_{\text{split}} : (N_{\text{proj}} - N_{\text{split}} + l^k)], \tag{2.47}$$

where $\boldsymbol{\iota}$ is a vector containing the equally spaced projection indices, l^k is a scalar index in the range $l \in [1, \ldots, N_{\text{split}}]$, and N_{split} is the number of projection divisions. This projection index may then select a subset as

$$\hat{\boldsymbol{y}}_j = [\boldsymbol{y}\{\boldsymbol{\iota}^k\}]_j \text{ for } j = 1, \ldots, \frac{N_{\text{proj}}}{N_{\text{split}}}, \tag{2.48}$$

where $\hat{\boldsymbol{y}}$ is the subset of the full set of measurements $\boldsymbol{y}$. It is usually assumed N_{split} is a factor of N_{proj}, so that $N_{\text{proj}} \mod N_{\text{split}} = 0$. With this, a simple consecutive ordered selection of the subsets can be performed as $l^k = (k \mod N_{\text{split}}) + 1$. Selection of l^k has been shown to have impact on the robustness of these types of approaches however [23], where one can adopt the 'bit reversal ordering' of [7]. This has the effect of reducing the correlation and error accumulation between consecutive iterations, and interestingly outperforms a random selection [23].

Two potential issues with subset gradient descent methods are that there will be some estimation bias between their estimate and the final reconstruction [115], they will not converge to a fixed point, and they may well diverge [23]. To address these points, Ahn and Fessler [116] introduced a method that progressively relaxes the step-size, which may be updated as

$$\lambda^{k+1} = \frac{\lambda^0}{\gamma(k-1)+1}, \tag{2.49}$$

where λ^0 is the initial step-size, and γ is a scalar factor to set the rate at which the steps diminish. The tendency to divergence may be reduced by increasing the size of the subsets, but this will ultimately limit the acceleration potential [23].

Momentum

The concept of momentum in the context of gradient descent methods, is supplementing the direction of travel with information from previous updates. This has the effect of aggressively increasing the rate of convergence.

An analogy to descent can be made through considering a ball rolling down a hill to the bottom of a valley. Polyak [117] used this physical interpretation to provide an acceleration strategy, where one considers the momentum of a heavy ball rolling down a hill, which causes it to descend very quickly. Applying Newton's laws to an object in a 1D gravitational potential well can be described as [22]

$$\ddot{x}(t) + a\dot{x}(x) + b\nabla f(x(t)) = 0, \tag{2.50}$$

where $a > 0$ is the friction and b is the gravitational force. In discrete variables and after

rearrangement this may be written as

$$x^{k+1} := x^k - \lambda \nabla f(x^k) + \beta(x^k - x^{k-1}), \tag{2.51}$$

where α is the step size and β is the term for the 'inertia'. More recently, the model has been extended to a proximal algorithm to incorporate nonsmooth regularisation as

$$\boldsymbol{\mu}^{k+1} := \mathbf{prox}_{\lambda^{-1}\mathcal{R}}\left(\boldsymbol{\mu}^k - \lambda \boldsymbol{\nabla} \mathcal{D}(\boldsymbol{\mu}^k, \boldsymbol{y}) + \beta\left(\boldsymbol{\mu}^k - \boldsymbol{\mu}^{k-1}\right)\right). \tag{2.52}$$

The algorithm in (2.52) is known as the inertial proximal algorithm for nonconvex optimisation (iPiano) [22], which as the name suggests, makes it applicable also for nonconvex objective functions such as the NLL in CT.

Another similar acceleration strategy was proposed by Nesterov [118], which was algebraically derived rather than motivated from physics as in the heavy ball. Probably the most popular instance of Nesterov's acceleration is known as the fast iterative shrinkage-thresholding algorithm for linear inverse problems (FISTA) [10], which is given as

$$\boldsymbol{\mu}^k := \mathbf{prox}_{\lambda^{-1}\mathcal{R}}\left(\boldsymbol{z}^k - \lambda \boldsymbol{\nabla} \mathcal{D}(\boldsymbol{z}^k, \boldsymbol{y})\right), \tag{2.53}$$

which is dependent on the intermediate parameter $\boldsymbol{z}$ updated as

$$\boldsymbol{z}^{k+1} := \boldsymbol{\mu}^k + \frac{t^k - 1}{t^{t+1}}(\boldsymbol{\mu}^k - \boldsymbol{\mu}^{k-1}), \tag{2.54}$$

with the factor t updated as

$$t^{k+1} := \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}. \tag{2.55}$$

When developing a fast reconstruction algorithm, one may wish to combine preconditioning, subset gradients and momentum. This was investigated by Kim et al. in [23] and later demonstrated by Wang et al. [119] to provide an acceleration factor of almost 50× over a full (diagonally preconditioned) gradient method for CBCT. This effectively reduced a reconstruction time from over an hour to within 2 minutes for real head data.

Hardware Acceleration

So far, we have discussed algorithmic acceleration strategies to improve the convergence rate of gradient based reconstruction. In practice, any algorithm will only be as fast as the hardware it is running on, so this is another critical consideration for fast techniques.

During reconstruction, presuming the regularisation function is cheap to compute, the brunt of the computational load will be in calculating the forward and backprojection operations $\mathbf{\Phi}$ and $\mathbf{\Phi}^T$. As linear operations, they can be subject to mass parallelisation, since each ray can be treated as independent. Particularly well suited hardware to this type of operation are graphics processing units (GPUs) that can nowadays contain thousands of parallel processing units — a recent review on GPU usage for CT reconstruction was conducted in [120]. Along with GPUs, other parallel architectures such as field programmable gate arrays (FPGAs) or cell broadband engine architecture (CBEA) have also been evaluated for CBCT reconstruction [121].

Of CBCT reconstruction methods reviewed, the FDK is likely be the most amenable to acceleration through GPU implementation, since the bulk of the computational load is in a single backprojection, that in theory can be performed entirely in parallel. Lesser et al. [122] reported a GPU FDK offering a 200× speed up over multi-core CPU implementation, for example. On the other hand, the iterative subset gradient descent approaches are inherently sequential, and complete parallelisation is not so straightforward [123]. Nevertheless, the authors in [124] have demonstrated TV regularised proximal gradient descent reconstruction within 2 minutes on GPU hardware, albeit from very few projections.

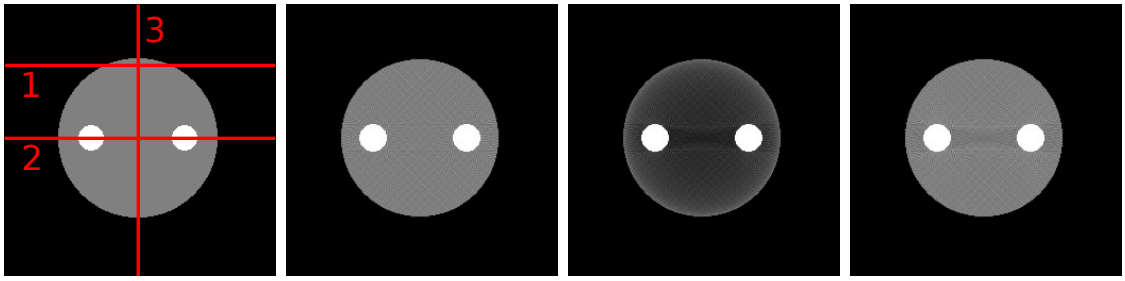
2.6 Mathematical Artefact Correction

In this section, we describe two common artefact correction methods: beam harding correction; and scatter compensation. We will expand on these ideas in Chapter 3 and Chapter 5 respectively, where we will use the techniques discussed here as bases for comparison.

2.6.1 Beam Hardening Correction

Beam hardening is an effect arising from the spectral dependency of X-ray attenuation. By observing the approximate parametrisation for photoelectric effect from (2.6), where its interactive cross section falls as ξ^{-3}, the lower energy photons are very easily attenuated along the trajectory. This in itself naturally leads to a beam hardening. The effect is however much more dramatic after materials with a high Z_{eff} such as bone or metal, where almost all of the low energy photons will be lost.

The consequence of the hardening of the beam is that the attenuation is nonlinear, and there are inconsistencies between different rays if this fact is ignored. To illustrate this effect, we have shown a simple numerical specimen in Figure 2.17a, corresponding of a circular region of water and two hard bone structures. From a set of linear parallel projections, we are able to faithfully reconstruct to a reasonable accuracy with FBP, which is shown in Figure 2.17b. The problem arises when the specimen is instead irradiated by polyenergetic radiation from a broad spectrum such is shown in Figure 2.18. One can see that through the different paths indicated in Figure 2.17a, there is a noticeable beam hardening effect in the spectra, especially through the bone inserts, where only the highest energy photons survive the trajectory. With these measurements, if the linear model is still enforced, then the resulting reconstruction from FBP in Figure 2.17c shows shading towards its centre and a dark streak between the two bone inserts — these are typical artefacts from the beam hardening effect [125].



(a) Numerical specimen (b) Monoenergetic FBP (c) Polyenergetic FBP (d) Corrected

Figure 2.17: Illustration of beam hardening effect for numerical phantom: (a) shows the numerical specimen with line profiles indicating rays for spectrum analysis; (b) shows reconstruction through FBP with a monoenergetic source; (c) is uncorrected reconstruction from a polyenergetic source; (d) is FBP reconstruction after LUT beam hardening correction.

A simple method for correcting the non-linearity of beam-hardening is by making the assumption that everything in the specimen has a mass attenuation coefficient of water.

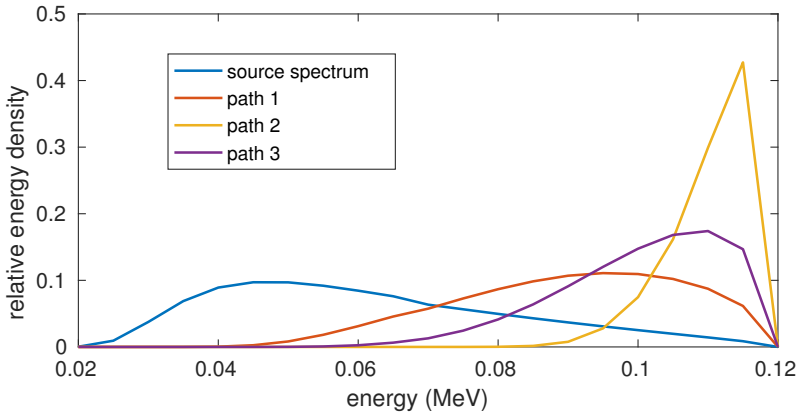


Figure 2.18: Detected spectra through paths indicated in Figure 2.17 to illustrate beam hardening through thick regions.

This may be done by writing the linearisation of a polyenergetic beam on water in the discretised setting as

$$\boldsymbol{p}_{\text{poly}} = \log\left(\boldsymbol{I}_0 \oslash \left[\sum_{j=1}^{N_\xi} \boldsymbol{b}(\xi_j) \odot \exp\left(-m_{\text{water}}(\xi_j)\boldsymbol{\Phi}\boldsymbol{\rho}\right)\right]\right), \tag{2.56}$$

where we define the total incident flux $\boldsymbol{I}_0$ as

$$\boldsymbol{I}_0 = \sum_{j=1}^{N_\xi} \boldsymbol{b}(\xi_j). \tag{2.57}$$

We would then like to calibrate this to a monoenergetic attenuation as

$$\boldsymbol{p}_{\text{mono}}(\xi_0) = \log\left(\boldsymbol{I}_0 \oslash \left[\boldsymbol{I}_0 \odot \exp\left(-m_{\text{water}}(\xi_0)\boldsymbol{\Phi}\boldsymbol{\rho}\right)\right]\right) = m_{\text{water}}(\xi_0)\boldsymbol{\Phi}\boldsymbol{\rho}. \tag{2.58}$$

By combining (2.58) with (2.56), one can rewrite this as

$$\boldsymbol{p}_{\text{poly}} = \boldsymbol{h}(\boldsymbol{p}_{\text{mono}}(\xi_0)), \tag{2.59}$$

with the function

$$\boldsymbol{h}(\boldsymbol{z}) = \log\left(\boldsymbol{I}_0 \oslash \left[\sum_{j=1}^{N_\xi} \boldsymbol{b}(\xi_j) \odot \exp\left(-\frac{m_{\text{water}}(\xi_j)}{m_{\text{water}}(\xi_0)}\boldsymbol{z}\right)\right]\right). \tag{2.60}$$

Standard beam hardening correction can then be seen as calculating this function's inverse $\boldsymbol{h}^{-1}(\boldsymbol{p}_{\text{poly}})$. This is usually done through means of forming a look up table [126] or by

polynomial fitting [45, 127].

We applied this approach for $\xi_0 = 60$ keV, and the resulting reconstruction is shown in Figure 2.17d. Clearly, this simple calibration has corrected the intensity error to some degree, but a streak artefact between the bony structures is still visible.

Another approach is to account for the effect of beam hardening through a polyenergetic statistical model. Two such techniques are Poly-SIR [46] and IMPACT [11], which use the water–bone and photoelectric–Compton models in (2.6) and (2.9) respectively. We will evaluate the effectiveness of these approaches against our developed Polyquant model in Chapter 3.

2.6.2 Scatter Correction

Scatter contamination, arising from deflected Compton and coherent scattered photons reaching the detector, can be a significant source of errors in CT. This leads to corruption, since the line-of-sight measurements are overestimated, ultimately leading to an underestimation in the attenuation or 'shading artefacts' [125].

The correction of scatter is often regarded in a two step process: estimation [128] and compensation [129]. For the compensation of scatter, a simplistic approach may be to subtract a given estimate from the measured intensities as

$$\boldsymbol{y}_{\text{SC}} = \boldsymbol{y} - \tilde{\boldsymbol{s}}, \tag{2.61}$$

where $\boldsymbol{y}_{\text{SC}} \in \mathbb{R}^{N_{\text{ray}}}$ are the corrected measurements and $\tilde{\boldsymbol{s}} \in \mathbb{R}^{N_{\text{ray}}}$ is the estimate of scatter. There are potential problems with this subtractive approach, especially when the estimate of scatter is larger that the measured intensity — leading to physically impossible negative values in $\boldsymbol{y}_{\text{SC}}$. This becomes particularly critical when these are later linearised to projections, as in many common approaches such as FDK or PWLS, where the logarithm taken will become undefined. One approach to overcome this is to employ a thresholding function on the subtractive estimate as [130]

$$\tilde{s}_i' = \begin{cases} 0 & \text{if } \tilde{s}_i \leq 0 \\ \tilde{s}_i & \text{if } 0 < \tilde{s}_i < \beta y_i \\ y_i + (\beta - 1) y_i \exp\left(\frac{\tilde{s}_i - \beta y_i}{(1-\beta) y_i}\right) & \text{if } \tilde{s}_i \geq \beta y_i, \end{cases} \tag{2.62}$$

where $\beta \in (0, 1)$ is a factor to set the level of thresholding, and can be set to be high when one is confident in the scatter accuracy. In [130], the authors experimentally use $\beta = 0.99$. The function in (2.62) can be seen as a soft thresholding function, which ensures that both the scatter remains non-negative and that $s_i < y_i \, \forall i$, whilst smoothing the transition for very high scatter.

Another common approach is to use a multiplicative factor rather than subtractive compensation. In this case, the measurements are instead corrected as [8, 131]

$$\boldsymbol{y}_{\text{SC}} = \boldsymbol{y}(\mathbf{1} - \boldsymbol{\gamma}), \tag{2.63}$$

where $\boldsymbol{\gamma}$ is a compensation factor that can be calculated as [8]

$$\gamma = \begin{cases} \beta & \text{if } \tilde{s}_i \geq \beta y_i \\ \frac{\tilde{s}_i}{y_i} & \text{otherwise} \end{cases} \tag{2.64}$$

where like in (2.62) $\beta \in (0, 1)$ is a thresholding factor, and the authors in [8] use $\beta = 0.95$.

It should be noted that when the gradient of the NLL is used directly for reconstruction as in (2.27) then compensation is not necessary, and one only needs to acquire an estimate $\tilde{\boldsymbol{s}}$. In any case, one requires some means to estimate the scatter, and we broadly classify this process as either 'image based' and 'convolutional based' — also known as scatter kernel superposition (SKS) [8].

Image Based Estimation

In image based estimation, one tries to physically model the scatter through a reconstructed volume. Needless to say, a dependency problem arises in generating this reconstruction, as the measurements are contaminated by the scatter we wish to estimate. Empirical iterative approaches have been proposed to partially mitigate this effect [129]. Despite this, given that one does possess a decent representation of the underlying object, then in theory one can analytically calculate the expectation of scatter, by evaluating the coherent and Compton scatter probabilities along all the possible scatter paths for a known source. In practice, such a detailed approach is computationally infeasible, and one must employ some degree of approximation to realise these methods.

One such approximation technique is known as Monte Carlo modelling, which is a concept we will explore again for measurement synthesis in Section 2.7 and dose calculation in Chapter 5 . In essence, one has a highly detailed model for X-ray interaction with matter, which is probabilistically sampled for a finite number of photons. An expectation can be estimated from these single instances of scatter, by combining them and then using some degree of smoothing. Examples of Monte Carlo scatter estimation implementations for CBCT are [131–134].

Another approach is to derive a practical analytic scatter model, by using strong approximations of the true scattering properties. For example, in [135] the authors derive a 'single scatter model', which models scatter as having only undergone a maximum of one interaction. A single scatter model can be formulated in a simple manner: by first projecting the source into the specimen for each gantry angle and recording the received X-ray intensity at each voxel location; then projecting from each voxel location onto the detector, weighted by the scatter probability for a given solid angle from the Klein–Nishina formula [40]. Despite this perceived simplicity, this is still a very expensive process, since the number of required projection operations is $N_{\text{proj}}(N_{\text{vox}}+1)$. Therefore, for its expense to be comparable to the cost of reconstruction, heavily sub-sampling both the object and detector are likely required. One will then incur losses from both neglecting multiple scatter events, and the loss of image structure through this sub-sampling.

Convolutional Scatter Estimation

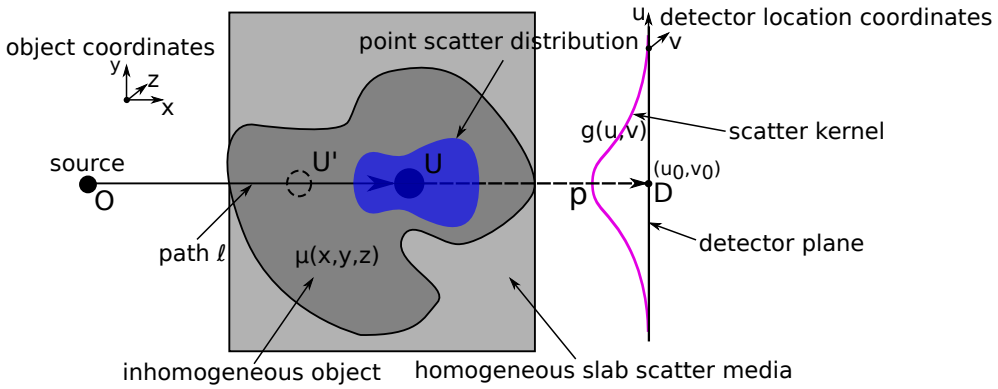


Figure 2.19: Scatter diagram showing homogeneous slab media and coordinate systems of object and detector along with point scatter distribution for forward scatter derivation.

A simple form of an SKS convolutional scatter model, as proposed by Ohnesorge et

al. [136], is represented in the discrete domain as

$$\boldsymbol{s}\{k\}(u,v) = \boldsymbol{p}\{k\}(u,v) * \boldsymbol{g}(u,v) \text{ for } k = 1, ..., N_{\text{proj}}, \tag{2.65}$$

where we use the notation $\{k\}(u,v)$ to indicate the k^{th} projection, with coordinates in the 2D detector plane (u,v) as in Figure 2.19, N_{proj} is the number of projections, and $*$ represents a 2D convolution across the (u,v) dimensions. With this, the scatter is modelled as some intensity function $\boldsymbol{p}$ filtered by a spatially invariant kernel $\boldsymbol{g}$. The assumption implied by this spatial invariance is that the scatter paths are through some homogeneous slab media, instead of the true inhomogeneous object, as is illustrated in Figure 2.19.

The scatter kernel function $\boldsymbol{g}$ in (2.65) often takes the form of the sum of two 2D Gaussian functions [8, 136], which offer a close empirical fit to real profiles, as

$$g(u,v) = a_{\mathcal{N}} \exp\left(-\frac{(u-u_0)^2 + (v-v_0)^2}{c_{\mathcal{N}}^2}\right) + a_{\mathcal{B}} \exp\left(-\frac{(u-u_0)^2 + (v-v_0)^2}{c_{\mathcal{B}}^2}\right), \tag{2.66}$$

where (u_0, v_0) are the coordinates at the centre of the detection plane, and a and c are the amplitude and width of the two Gaussian functions. We use the subscripts $\mathcal{N}$ and $\mathcal{B}$ to indicate the narrow and broad functions respectively.

The scatter intensity function $\boldsymbol{p}$ in (2.65) has some physical meaning, as it is the amount of scattered photons arriving along the initial trajectory of the ray. With reference to Figure 2.19, the derivation of this factor according to [136] may be made in the continuous domain for a monoenergetic beam and ignoring multiple scattering. To begin with, the photon flux at point U after travelling through the object is expressed as

$$I_U = I_O \exp\left(-\int_O^U \mu(\ell)\, d\ell\right). \tag{2.67}$$

At this point, photons will have a probability of being scattered at any angle by electrons in the object. We will denote the proportion of photons scattered at an angle of 0° as $K\mu(U)$, which is assumed proportional to the attenuation coefficient at that point. In reality, there will be a dependence on the spectrum of the source and the electron density, and we will explore this when developing the PolySKS model in Section 4.2.1. For this to be detected at the point D, with coordinates (u_0, v_0) in this case, it must travel through

the rest of the object. The expectation of scatter from this single point is therefore

$$I_D = I_U K \mu(U) \exp\left(-\int_U^D \mu(\ell)\, d\ell\right). \tag{2.68}$$

The total scatter factor is the summation of all the individual scatter factors from O to D, which can be expressed as

$$\begin{aligned} p &= \int_O^D I_O \exp\left(-\int_O^{U'} \mu(\ell)\, d\ell\right) K\mu(U') \exp\left(-\int_{U'}^D \mu(\ell)\, d\ell\right) dU' \\ &= I_O \exp\left(-\int_O^D \mu(\ell)\, d\ell\right) K \int_O^D \mu(\ell)\, d\ell. \end{aligned} \tag{2.69}$$

By now adopting discrete notation, this may be expressed with respect to the system matrix $\boldsymbol{\Phi}$ as

$$p_i = K I_{\text{in}} \exp(-[\boldsymbol{\Phi}\boldsymbol{\mu}]_i)[\boldsymbol{\Phi}\boldsymbol{\mu}]_i \text{ for } i = 1, ..., N_{\text{ray}}, \tag{2.70}$$

where $I_{\text{in}} = \sum_{j=1}^{N_\xi} b(\xi)$ is the total incident flux of the source. Now with (2.70), it is possible to calculate the scatter factor at each iteration according to this monoenergetic model, and we will explore this concept with the Int-fASKS method tested in the experimental section of Chapter 4. In practice however, one often performs the estimation as a preprocessing operation before reconstruction, when one has no knowledge of $\boldsymbol{\mu}$. By assuming the linearity from a monoenergetic source, an estimate is possible through calculating

$$p_i = K y_i^{h_1} \left[-\log\left(\frac{y_i}{I_{\text{in}}}\right)\right]^{h_2} \text{ for } i = 1, ..., N_{\text{ray}}, \tag{2.71}$$

where the parameters h_1 and h_2 are heuristic and provide a better fit to a real source, and selected to match observation [136]. They may be fitted by using a non-linear least squares approach such as Newton's method to experimental or simulated kernels. The effect of these parameters may compensate for some of the approximations including: the monoenergetic model, and only single scattering events from the forward scatter model.

Along with the before-mentioned approximations of this model, there are a couple of limitations in calculating forward scatter using (2.71) instead of (2.70). Firstly, it is calculated using the physical measurements $\boldsymbol{y}$, which will themselves be contaminated by the scatter one is attempting to estimate, and will lead to a biased estimate. In an attempt to mitigate this, scatter estimation may be performed iteratively, where $\boldsymbol{y}$ is first corrected for scatter, then used as the basis for the next estimate as in [8]. This process then acts as

a deconvolution, which is known to be non-exact with the presence of noise. This iterative correction also presents a second problem in cases when the scatter expectation is larger that the measurements themselves, which is possible with CBCT [137], and where the $\log(\cdot)$ will be undefined. Applying thresholding to the corrected $\boldsymbol{s}$ to ensure it remains positive as in [8, 60] will allow the calculation of (2.71), but at the expense of introducing a bias.

Due to the high dimensionality of the CBCT measurements, one can perform the convolution in (2.65) efficiently through use of the FFT. With this, the order of complexity can be reduced from $\mathcal{O}(N_{\text{proj}}N_u^2N_v^2)$ to $\mathcal{O}(N_{\text{proj}}N_uN_v\log(N_uN_v))$ for a detector with dimensions $N_u \times N_v$. The re-expression of (2.65) employing FFT operations is

$$\boldsymbol{s}\{k\}(u,v) = \mathcal{F}^{-1}\left[\mathcal{F}\left(\boldsymbol{p}\{k\}(u,v)\right) \odot \mathcal{F}\left(\boldsymbol{g}(u,v)\right)\right] \text{ for } i = 1, ..., N_{\text{proj}}, \tag{2.72}$$

where $\mathcal{F}[\cdot]$ and $\mathcal{F}^{-1}[\cdot]$ are FFT and inverse FFT (IFFT) operations. To avoid aliasing and to allow an efficient FFT decomposition, some degree of zero padding is necessary, which may typically double the effective number of samples.

It was demonstrated in [8] that even for homogeneous slabs of water, both the parameters of the forward scatter calculation K, h_1 and h_2, and the width of the Gaussian kernels $c_{\mathcal{N}}$ and $c_{\mathcal{B}}$ are dependent on the thickness of the slab for a typical energy spectrum. To account for this and still enable fast convolution, they decomposed the projection into several thickness groups, each with a different set of parameters. On top of this, they also include a factor to account for inhomogeneities at edges, and linearly modulate the kernel amplitude with the object thickness to approximate more realistic scatter media. In combination, this is known as the fast adaptive SKS (fASKS) and can be expressed as [8]

$$\begin{aligned}
&\boldsymbol{s}\{k\}(u,v) = \\
&(1 - \gamma\boldsymbol{\tau}\{k\}(u,v)) \odot \mathcal{F}^{-1}\left[\sum_j^{N_{\mathcal{R}}} \mathcal{F}\left(\boldsymbol{\mathcal{R}}_j\{k\}(u,v) \odot \boldsymbol{c}_j\{k\}(u,v) \odot \boldsymbol{p}_j\{k\}(u,v)\right) \odot \mathcal{F}\left(\boldsymbol{g}_j(u,v)\right)\right] \\
&+ \gamma\mathcal{F}^{-1}\left[\sum_j^{N_{\mathcal{R}}} \mathcal{F}\left(\boldsymbol{\tau}\{k\}(u,v) \odot \boldsymbol{\mathcal{R}}_j\{k\}(u,v) \odot \boldsymbol{c}_j\{k\}(u,v) \odot \boldsymbol{p}_j\{k\}(u,v)\right) \odot \mathcal{F}\left(\boldsymbol{g}_j(u,v)\right)\right]
\end{aligned} \tag{2.73}$$

where $N_{\mathcal{R}}$ are the number of thickness groups, $\boldsymbol{\mathcal{R}}_j\{k\}(u,v)$ are binary masking functions to isolate the projection elements belonging to the appropriate thickness interval, $\boldsymbol{c}_j\{k\}(u,v)$ are factors to linearly attenuate the scatter kernels close to edges, γ is a scalar constant to weight the strength of the thickness modulation, and $\boldsymbol{\tau}\{k\}(u,v)$ is the water equivalent thickness of the projection given as

$$\boldsymbol{\tau}\{k\}(u,v) = -\frac{1}{m_w}\log\left(\boldsymbol{y}\{k\}(u,v) \oslash \boldsymbol{I}_{\text{in}}\{k\}(u,v)\right) \text{ for } i = 1, ..., N_{\text{proj}}, \tag{2.74}$$

where m_w is the mass attenuation coefficient of water at some nominal photon energy, and $\oslash$ denotes the element-wise division.

Due to the adoption of several heuristics in order to get a convolutional scatter model to perform well in practice, it leads one to wonder if a transformation could instead be somehow 'learnt' from a set of scatter images. A similar concept was investigated as early as 1993 [138], using an artificial neural network for estimating scatter from chest scans. More recently, deep convolutional neural networks have been applied to the forward scatter factor in (2.70), and may offer promising performance [139].

2.7 Accurate Computer Simulation of X-ray Computed Tomography

Although CT scanners are physical tools, and any reconstruction method should ultimately be evaluated through physical experiment, one seldom has ground truth information against which to compare their performance. An exception to this is with synthetic phantoms, of which one usually has very well defined structural and chemical information to form the basis of a gold standard. To evaluate more complex structures, generate auxiliary information not available from real scanners, and generate novel or extreme scanning situations, simulation tools are valuable.

In this section, we describe the simulation of CT systems, and how this can be used for low-dose reconstruction testing. As well as general modelling considerations, we will detail the specific combination of software tools, which we have used to generate synthetic CT data for numerical results throughout this thesis. At its core, this is using the Gate Monte Carlo software [26]. This approach was presented in [140].

2.7.1 Monte Carlo Simulation of X-ray Interaction

Monte Carlo simulators allow accurate physical modelling of the X-ray interactions with matter. They work by propagating some number of photons through a discretised specimen, and probabilistically modelling the interactions along their path.

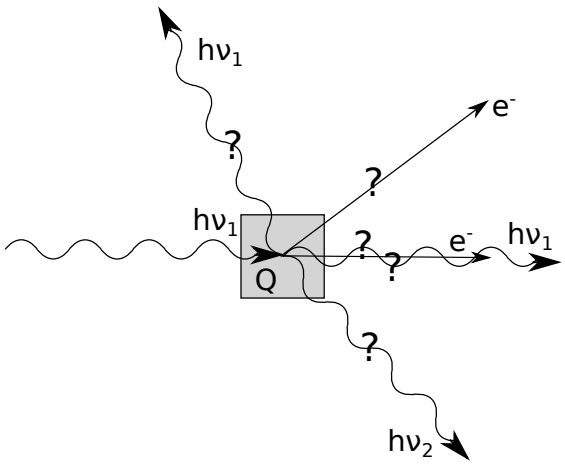


Figure 2.20: Illustration of probabilistic interaction event modelled through physical simulation.

To illustrate this concept, Figure 2.20 shows an X-ray photon of energy $h\nu_1$ entering a volume Q. At this point certain interaction events may occur, or the photon may go undisturbed and continue on its trajectory. For the energies in the spectrum of a typical X-ray source, these interactions are: photoelectric effect and the ejection of an electron; Compton scattering leading to a change in photon energy to $h\nu_2$ and ejection of an electron; or elastic scattering. The probability of any given interaction to occur will depend on the interaction cross section, which are well documented for all relevant elements and energies such as in [141]. One then requires a pseudo random number generator such as the Mersenne Twister [142], to select an interactive instance.

After the interaction at a given site, the photon is then tracked as it propagates through the rest of the specimen, before finally being deflected into empty space, attenuating within the specimen or hitting the detector, where it contributes to a measurement.

2.7.2 Modelling the X-ray Source

With a general method for simulating the interaction of any particles, one must then construct realistic elements as part of this simulation. Firstly, a realistic X-ray source is

critical for getting the desired interactions and responses. The approach we take for this is to mimic the spatial and spectral profile of a commercial source, though without having to model the interactions leading to these, to increase the efficiency.

To derive a simulated source, we generate an appropriate spectrum using an analytic approximation to a typical diagnostic X-ray source with software package SpekCalc [143]. Along with modelling the expected spectrum from the bremsstrahlung and characteristic radiation of an angled tungsten anode as shown in Figure 2.3, this also includes a titanium beam hardening filter and aluminium equivalent attenuation of the window and other components in the X-ray path, all according to the specifications of a commercial CBCT system as illustrated in Figure 2.11.

Another important feature to model is the flux resulting from the bow-tie filters, shown in Figure 2.12. With this, not only does one need a spatial variation in intensity but also in spectrum. To model this element, we made an approximation amounting to taking the mean thickness of bow-tie allowing for a single spectrum. The flux however from a commercial system was replicated by taking the flat field image from a scanner, and setting this as the source intensity.

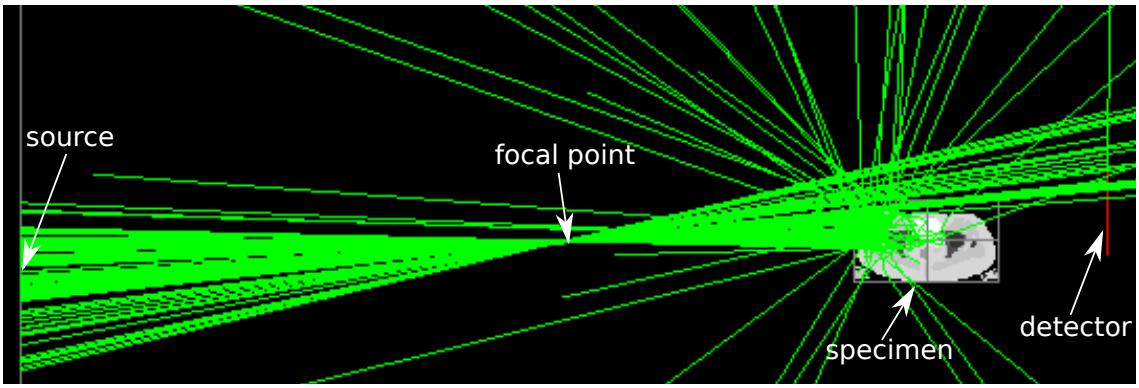


Figure 2.21: Illustration of source position and focal point to replicate beam profile.

To model a divergent source for the CBCT, we mirrored the fluence image and set it on the opposite side of a single focus point. This has the effect of reproducing the desired intensity at the detector — this method is illustrated in Figure 2.21.

2.7.3 Modelling the Detector and Collimator

A cross section of a modelled detector with focused collimator in shown in Figure 2.22. In this illustration, several photon trajectories are shown whose direction is from top to

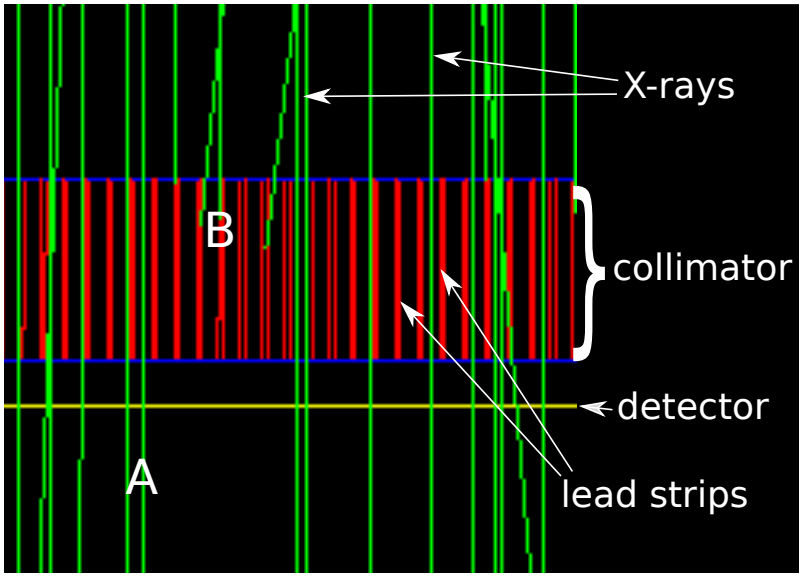


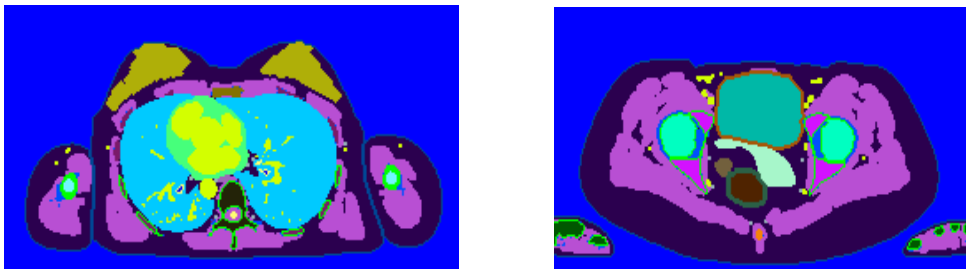
Figure 2.22: Illustration of physical collimator object with simulated photon events shown in green.

bottom. With reference to the diagram, one can see how trajectories parallel to the strips such as those in region A pass through the detection plane where they are accumulated, and photons from scattered trajectories such as those in region B are attenuated by the collimator.

2.7.4 Modelling the Specimen and its Rotation

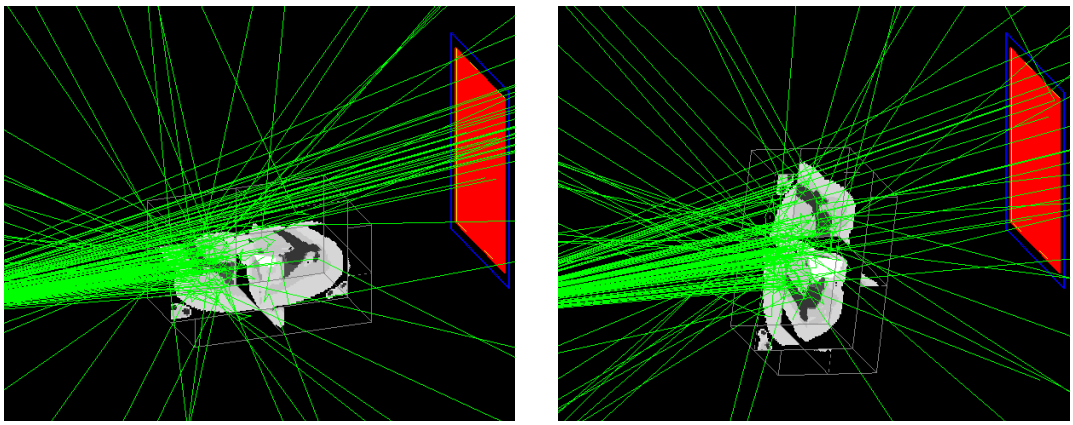
The interaction of X-ray radiation with the specimen of interest is characterised with the attenuation coefficient $\boldsymbol{\mu}(\xi)$ for a polyenergetic source. In order to model the interactions through Monte Carlo simulation, one requires the chemical composition and density of these materials, which in turn will characterise the attenuation. One such data set is the adult reference computational phantom (ARCP) [41], of which a slice of the chest is shown in Figure 2.23a, along with the pelvis in Figure 2.23b. In this work, we are concerned with the electron densities, which can be analytically calculated from the chemical properties according to (2.4), providing us with ground-truth information.

The acquisition of CT data involves the rotation of the source and detector around the specimen, whilst collecting projections. For simulation, we perform this process by instead rotating the specimen by discrete angles after a set number of photons. Examples of this along with an illustration of X-ray flight paths for 100 photons is illustrated in Figures 2.24a and 2.24b.



(a) Coloured phantom of chest (b) Coloured phantom of pelvis

Figure 2.23: Numerical specimens: (a) is a segmented material class map for polyenergetic simulation, with materials in arbitrary colours; (b) is the pelvis region.



(a) 0° acquisition (b) 72° acquisition

Figure 2.24: Illustration of simulation of CBCT system for two different projection angles, for 100 photons (green) in each case.

2.7.5 Computational Implementation and Usage

Realising a computationally efficient implementation of a Monte Carlo simulation is critical for making it feasible. One convenient feature is that each photon track may be considered as independent, allowing the calculation to be split into many parts, before accumulating the result to form the complete system. This makes the calculation available for parallel hardware such as a CPU cluster, or GPU.

An example script to perform CBCT simulation according to the method described in this section is given in Appendix B. This is a complete simulation, which should be split and ran in parallel.

In our implementation, we utilised the Eddie Mark 3 cluster at the University of Edinburgh's Compute and Data Facility. This is a shared computational resource consisting in over 7000 Intel Xeon cores. In order to utilise the CPU cluster, we executed

subsets of the simulated particles on separate machines. Doing this splitting of a simulation is functionality now built into the Gate software [26]. We then concatenate all the separate simulations in MATLAB ahead of its use in reconstruction.

In Chapter 3, we use this simulation approach to generate the numerical fan-beam chest and metal hip implant data. We also use this tool extensively for scatter modelling for Chapter 4. Firstly, we simulate the scattering through water slabs in order to fit and analyse our estimation method. We also synthesise CBCT head of head and pelvis regions in the numerical reconstruction section. Finally, we then reuse this CBCT for the dose calculation validation study in Chapter 5.

Chapter 3

Direct Quantitative Reconstruction

3.1 Introduction

Physically quantifying reconstructions from CT is highly valuable in some medical practices. In radiation therapy for example, as is the focus of this work, the electron density inferred from CT images of the patient allows the dose deposition from the treatment beam to be accurately modelled [144]. Additionally, quantifying the bone mineral density allows osteoporosis to be characterised and the risk of bone fractures to be assessed [145]. Due to the non-linear energy-dependence of X-ray attenuation however, mapping from a set of raw measurements to a consistent physically quantitative reconstruction is not straight forward, and requires both actively accounting for the polyenergetic effects such as beam hardening, and establishing a relation to the quantity of interest.

Mapping from CT to physical density is usually treated as a several step process: linearising the measurements by correcting for scatter, as covered in Section 2.6.2; taking the logarithm [60] and non-linearly calibrating from a polyenergetic to approximate monoenergetic source [45], as covered in Section 2.6; inverting the linearised projections with analytic or iterative reconstruction algorithms [55], as covered in Section 2.4 and Section 2.5 respectively; then finally applying a non-linear calibration to mass or electron density [30], using calibration curves shown in Figure 2.8b.

In general, the mapping from a single polyenergetic to a monoenergetic measurement is an ill-posed problem, since the spectral information of the irradiated material is lost with the standard energy integrating detectors [37]. Whilst imaging the human body however, most tissues may be reasonably modelled with few distinct materials: namely water and bone as in [45], which is shown in (2.9). If the amount of each material in a projection is known a priori from a water–bone segmentation, then it is possible to approximately reverse the polyenergetic effects through polynomial fitting or with a look up table. For more

accuracy, this may also be brought into the reconstruction model, where [46] demonstrate statistical quantitative imaging of mass density, with the water–bone assumption. Given that the composition is itself highly correlated with density, then the prior segmentation may not be necessary, but estimated during reconstruction [47].

Another approach is to model the attenuation explicitly in terms of physical processes, given quantitative physical parameters of interest. One such choice is to model the photoelectric and Compton effects in terms of relative atomic number and electron density as shown in (2.6), which is reasonably accurate for relevant elements and energies [39, 43, 44]. Given measurements from two sufficiently different spectra—a technique known as dual-energy CT (DECT)—a projection of Compton attenuation can be uniquely determined [42], of which electron density is an analytic function [40]. This DECT technique effectively bypasses the need for assumptions such as water–bone compositions, so should be applicable to a wider range in materials, although one faces practical difficulties in generating the two spectra [146]. Additionally, one will expect a loss in accuracy between significantly different elemental compositions, such as soft tissues, bone and metallic implants, since the two parameter model is not consistent over a wide range in atomic species [39].

In [11], the authors introduce an iterative maximum-likelihood polychromatic algorithm for CT (IMPACT), which models the energy independent factors in the photoelectric–Compton model from [42] as piecewise-linear functions of monoenergetic attenuation, allowing reconstruction from a single source. This method does not require any prior segmentation, and is reported on a wide range of materials including metallic implants. By imposing the energy dependence of the two parameter model however, the physical consistency will also degrade throughout diverse material types, due to the inconsistency of these parameterisations in effective atomic number and electron density [39].

The second conversion from either reconstructed attenuation in Hounsfield units (HU) in [11, 45] or mass density in [46, 47] to electron density is also an ill-posed problem. This is because X-ray interaction depends on the environment of the electrons as well as their density, and will vary considerably with atomic number for imaging energy ranges [37, 39]. Although, again given the fact that most human tissues have similar properties, a single piecewise linear fitting is reasonably accurate in practice [147], though it will not be consistent with synthetic materials such as some plastics [30]. Typical attenuation to

density calibration curves are shown in Figure 2.7.

In both the photoelectric–Compton and a material decomposition such as the water–bone model, there is some degree of fitting to materials and energies of interest. Instead, one could use a purely data-centric approach. Here, given a representative set of substances, one could use a model that accurately parameterises the energy dependent attenuation in terms of the quantity of interest, without necessarily any physical justification. One such method is to model the energy dependent attenuation as a piecewise-linear function of quantitative density, which may be fit to a set of materials of interest. When the transitions between linear sections are independent of energy we also get the nice property that the computation order in an iterative method is independent of the number of energies considered.

In this chapter, we will study in detail the piecewise linear quantitative model for CT, and will show how it may be incorporated in a regularised iterative reconstruction algorithm. Specifically, this provides a generalised method for directly quantifying the electron or mass density of a heterogeneous specimen, and is also able to model hard metallic structures without any prior segmentation. In [148] we presented a preliminary study using this idea for the specific case of calculating electron density for radiotherapy planning. The complete method, as presented in this chapter, was then published in [149].

During the development of this work, we became aware of the commercial method DirectDensity from Siemens Healthineers [150], which reports direct reconstruction into relative electron density from a single polyenergetic source. This is a preprocessing technique combining bone detection with a projection-based material decomposition similar to [45]. In this article we compare against the water–bone model of [46], which has been shown to itself provide superior performance to that of [45].

3.2 General Model for Polyenergetic Parameterisation

We propose to take a data centric rather than a physical approach to parameterising the X-ray attenuation. To motivate our choice for this, we have plotted the relative attenuation coefficient against electron and mass density for 52 biological tissues in [3] at a number of energies in Figure 3.1. We have normalised the attenuation to the maximum for an energy—the tooth in each case—simply to allow visualisation on a single graph.

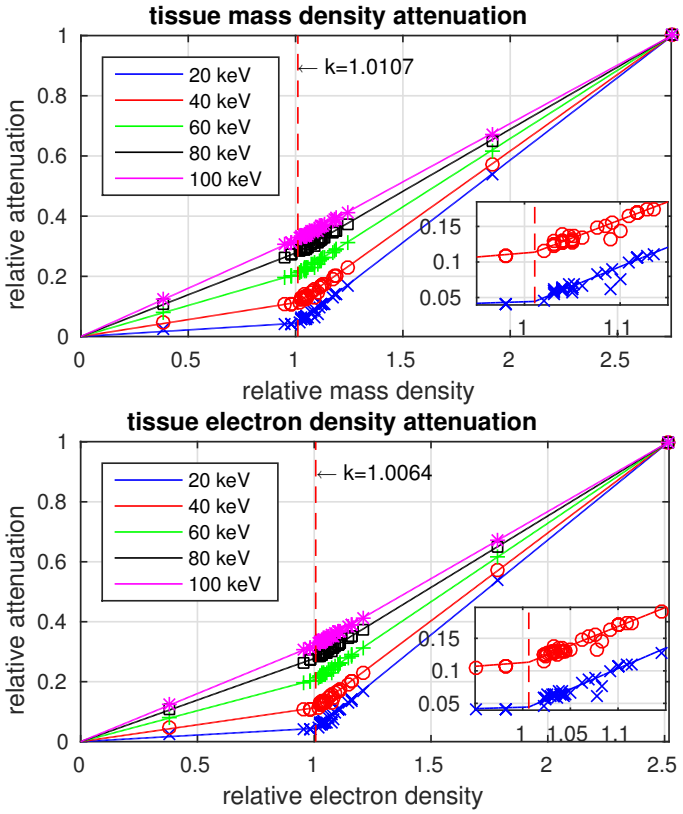


Figure 3.1: Relation between energy dependent attenuation and relative electron and mass density for biological tissues from [3]. The attenuation in each case is normalised to that of tooth, to allow visualisation on a single graph.

It appears from Figure 3.1 that the chosen biological tissues approximately follow an energy dependent two part linear fit in both electron and mass density. We have superimposed such a fit onto each plot in Figure 3.1, where the transition point k is fixed to be the same for each energy and set to minimise the norm of the residual through all the data. Imposing a constant k throughout will motivate computational efficiency outlined in the following section. One interpretation of this parameterisation is taking the inverse of HU to density shown in Figure 2.7 and incorporating a dependency on energy.

A key consideration of these observed trends are that they are purely empirical and do not necessarily have connection to a physical model. Additionally, it is not clear how this would have any connection to the approximate parameterisation in (2.6). This can be demonstrated by re-writing (2.6) as

$$\mu(\xi) = (K_1 Z_{\text{eff}}^{2.94} \xi^{-3} + K_2 f_{\text{KN}}(\xi))\rho_e. \tag{3.1}$$

Since there are no constant terms in (3.1), it follows that the zero offset in the linear trend

after the knee position in Figure 3.1 must have some non-trivial dependence on Z_{eff} is this model were to be used. Since it has been shown, however, that (3.1) is indeed not accurate for diverse materials [43] [39], we will not attempt to make this connection.

Generalising the piecewise linear fit motivated from the data in Figure 2.8b may be written as

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}, \xi) = \sum_{i=l}^{N_f} \boldsymbol{f}_i \odot (\alpha_l(\xi)\boldsymbol{x} + \beta_l(\xi)), \tag{3.2}$$

where N_f is the number of linear fits, f_i is a class indicator function for materials belonging to the l^{th} class, and $\boldsymbol{x} \in \mathbb{R}^{N_{\text{vox}}}$ is the physical parameter one wishes to fit to—electron density or mass density for example. The class indicator functions are defined as

$$f_{l,i} = \begin{cases} 1 & \text{if } k_{l-1} \leq \rho_{e,i} < k_l \\ 0 & \text{otherwise} \end{cases} \quad \text{for } l = 1, \ldots, N_f; \text{ for } i = 1, \ldots, N_{\text{ray}}, \tag{3.3}$$

where k_i is the 'knee' between the l^{th} and $(l+1)^{th}$ linear fits, with $k_0 = 0$. To ensure connected fits, we may also enforce the constraints

$$\alpha_l k_l + \beta_l = \alpha_{l+1} k_l + \beta_{l+1}. \tag{3.4}$$

We also have $k_{N_f} = 0$ and $\beta_1 = 0$, which ensures that a vacuum has no attenuation.

Following on from the connection to the decomposed inverse of calibration in Figure 2.7, this general model would indeed allow several breakpoints around different tissue classes as in [30], but here we focus on two or three for computational simplicity.

Special Cases and Connections to Existing Models

We now give several interesting special cases and connections to existing models:

- An example of (3.2) may be used for parameterising the attenuation of biological tissues as a function of relative electron density, which was the singular case studied in [148], and allows the fitting in Figure 3.1. For this we have $N_f = 2$, $x = \rho_e$ and $k = 1.0064$, which may be written as

$$\hat{\boldsymbol{\mu}}(\boldsymbol{\rho_e}, \xi) = \boldsymbol{f}_1(\boldsymbol{\rho_e}) \odot \alpha_1(\xi)\boldsymbol{\rho_e} + \boldsymbol{f}_2(\boldsymbol{\rho_e}) \odot [\alpha_2(\xi)\boldsymbol{\rho_e} + \beta_2(\xi)]. \tag{3.5}$$

This is the instance we study, along with an equivalent fit for mass density.

- The model above may also include highly dense materials, such as titanium metallic implants with a mass density of $\rho = 4.5$ g/cm^3. To include this, a second knee point may be inserted between bone and metal, and one would have $N_f = 3$. We also study this in our numerical experiment with hip implants.

- Another interesting special case with (3.2) is when all b_i are set to 0, $x = \rho$, $N_f = 2$, $\alpha_1(\xi)$ and $\alpha_2(\xi)$ are mass attenuation coefficients for water and bone, and the continuity constraints are ignored. This yields (2.9), with a density dependent threshold, and is equivalent to the formulation in [47] without smoothing. The consequence of ignoring the continuity in (3.4) will be a step at the 'knee' points.

- In a similar manner to the water–bone model utilised in [47], the IMPACT model in [11] may also be expressed as a special case of our general fitting model in (3.2). We note that adopting the same notation, the model may be written as

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}, \xi) = \sum_{i=1}^{N_f} \boldsymbol{f}_i \odot \left(\left[\frac{\alpha_i}{\xi^3} + \gamma_i f_{\mathrm{KN}}(\xi) \right] \boldsymbol{x} + \frac{\beta_i}{\xi^3} + \delta_i f_{\mathrm{KN}}(\xi) \right), \tag{3.6}$$

  where α_i, β_i, γ_i and δ_i are the coefficients or photoelectric and Compton linear fits, and $\boldsymbol{x}$ is a monoenergetic equivalent attenuation. A key difference between (3.6) and (3.2) is the energy dependence of coefficients and number of parameters. Whilst our model has $2N_fN_\xi - 1$ free parameters, (3.6) has $4N_f$, which is likely to be less, therefore our model offers increased model flexibility but, as we will see in the next section, at no additional computational cost.

- We could also use (3.2) to predict the reconstruction of a monoenergetic equivalent image—in DECT this is known as 'quasi monoenergetic' reconstruction [146]. This is possible since the imposition of a constant 'knee' position across energies allows any monoenergetic attenuation to be predicted using an equivalent piecewise linear function. The relation to the IMPACT model in this case is that the energy dependence is fitted from the data instead of implied from the photoelectric–Compton model in (2.6).

- Although not evaluated in this chapter, since the shape of HU to proton stopping power ratio (SPR) in [30] is extremely similar to the relative electron density curve observed in Figure 2.7, we suggest our model should be very applicable for SPR

calibration also.

3.3 Polyquant Method for Density Reconstruction

The Polyquant model introduced in Section 3.2 describes the forward mapping from physical quantity to attenuation. We now show how this may be combined with the statistical CT measurement model in (2.13), to allow direct statistical inference of mass or electron density.

Combining (3.2) with (2.13) results in the relation

$$\sum_{j=1}^{N_\xi} b_i(\xi_j) \exp\left(-[\boldsymbol{\Phi}\boldsymbol{\mu}(\xi_j)]_i\right) = \sum_{j=1}^{N_\xi} b_i(\xi_j) \exp\left(-[\boldsymbol{\Phi}\hat{\boldsymbol{\mu}}(\boldsymbol{x},\xi_j)]_i\right) \text{ for } i = 1, ..., N_{\text{ray}}. \tag{3.7}$$

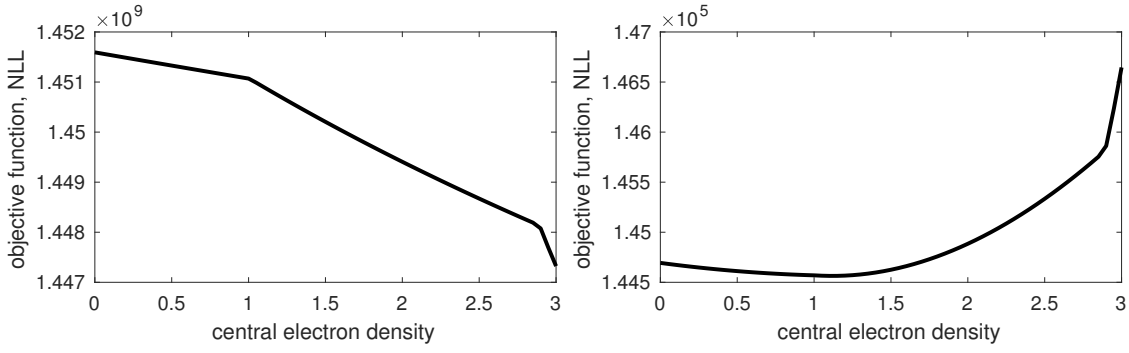
If we introduce a function $\psi(\cdot,\cdot)$ to simplify notation as

$$\psi_i(\boldsymbol{x},\xi) \equiv b_i(\xi) \exp\left(-[\boldsymbol{\Phi}\hat{\boldsymbol{\mu}}(\boldsymbol{x},\xi)]_i\right) \text{ for } i = 1, ..., N_{\text{ray}}, \tag{3.8}$$

we can write the negative log-likelihood (NLL) for the Poisson model as

$$\text{NLL}(\boldsymbol{x};\boldsymbol{y}) = \sum_{i=1}^{N_{\text{ray}}} \sum_{j=1}^{N_\xi} \psi_i(\boldsymbol{x},\xi_j) + s_i - y_i \log\left(\sum_{j=1}^{N_\xi} \psi_i(\boldsymbol{x},\xi_j) + s_i\right), \tag{3.9}$$

where we note that this function is non-convex as with similar CT NLL functions in [60,86].



(a) NLL around $\mathbf{0}$ (b) NLL around oracle

Figure 3.2: Objective function against central electron density value for chest data in Section 3.4.2: (a) is the case where all other pixels are set to 0; (b) is the case where all other pixels are set to the oracle values.

The demonstrate this non-convexity, we have included Figure 3.2. These were generated

by calculating the NLL as in (3.9), and varying the electron density for the central pixel. In the first case in Figure 3.2a, all other pixels are set to 0. Here, discontinuities at the knee positions are clear, with the second discontinuity due to a third piecewise linear interval for metal. Additionally, since the function will lie above a chord between points on different linear sections, this demonstrates non-convexity. On the other hand, around the oracle point, the NLL does appear to be both convex in this region, and minimised around the true value of $\rho_e = 1.1$. Due to this, care must be taken in minimising 3.9.

Reconstruction of the quantitative density map can be performed by finding an $\boldsymbol{x}$ that minimises (3.9). We will look at gradient descent methods, for which we require an expression for the derivative of NLL. If we simplify notation with the following

$$\boldsymbol{d}(\boldsymbol{x}) = \boldsymbol{y} \oslash \left(\sum_{j=1}^{N_\xi} \boldsymbol{\psi}(\boldsymbol{x}, \xi_j) + \boldsymbol{s} \right) - \mathbf{1}, \tag{3.10}$$

where $\oslash$ represents component-wise division. An expression for the derivative is then

$$\boldsymbol{\nabla} \mathrm{NLL}(\boldsymbol{x}; \boldsymbol{y}) \approx \sum_{i=1}^{N_f} \boldsymbol{f}_i(\boldsymbol{x}) \odot \boldsymbol{\Phi}^T \left[\sum_{j=1}^{N_\xi} \alpha_i(\xi_j) \boldsymbol{\psi}(\boldsymbol{x}, \xi_j) \odot \boldsymbol{d}(\boldsymbol{x}) \right], \tag{3.11}$$

where $\boldsymbol{\Phi}^T$ represents a transpose of the system matrix or 'back-projection', and $\odot$ is component-wise multiplication. We have shown this derivative as an approximation "$\approx$" in testament to the fact that there are discontinuities in the gradient at the 'knee' positions, where the gradient is not defined. Although in [148], we invoked the logistic function to mitigate this effect, we have found better empirical performance by instead using (3.11) directly.

Before we proceed, we note that (3.11) only has a single backprojection operation $\boldsymbol{\Phi}^T$ per linear fit, so the number of these operations is independent of the number of energies N_ξ. This is only possible due to constant 'knee' positions k for each energy. Calculating $\boldsymbol{\Phi}\hat{\boldsymbol{\mu}}(\boldsymbol{x}, \xi)$ is also independent on N_ξ, and can be evaluated with $2N_f - 1$ applications of $\boldsymbol{\Phi}$, where one fewer is a consequence of the assumption that $\beta_1 = 0$. For example, applying the method for electron density reconstruction of tissues using (3.5), one would expect a computation cost of three forward- and two backprojections.

Although one may obtain a quantitative density reconstruction through maximum likelihood estimation, by iteratively minimising (3.9) through gradient descent with

(3.11), incorporation of prior regularisation will typically improve statistical performance, especially as the noise increases or when few measurements are taken. Incorporating regularisation now gives the penalised log-likelihood or maximum a posteriori estimate as

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathcal{C}}{\operatorname{argmin}}\, \mathrm{NLL}(\boldsymbol{x};\boldsymbol{y}) + \lambda R(\boldsymbol{x}), \tag{3.12}$$

where $R(\cdot)$ is some regularisation function, and $\mathcal{C}$ is a set of box constraints on $\boldsymbol{x}$ so that $0 \leq x_i \leq \zeta$ for $i = 1, ..., N_{\text{vox.}}$, where ζ is the maximum allowable density value, and the constraint set ensures non-negative density values.

The choice of $R(\cdot)$ in (3.12) will vary on the imaging application, but some possibilities are generalised Gaussian Markov random field (GGMRF) [151], TV [98] or wavelet sparsity [152]. We opt in our experimental section for TV, since it promotes piecewise flat images, which we expect from homogeneous slabs of tissue.

Algorithm Design

Although there are many approaches for evaluating (3.12), we will detail here the algorithm we have used in our numerical experiments, which we give in Algorithm 1. In Section 3.5, we will present and analyse accelerated variants of this, which will become valuable later for real CBCT data experiments.

Algorithm 1 Prox-Polyquant

Initialization: Regularisation constant λ, Lipschitz approximation L_0, number of iterations N_{iter}, inertia factor $\gamma \in [0, 1)$, and starting point $\boldsymbol{x}^0$.
$\boldsymbol{x}^0 \leftarrow \mathbf{1}$
$\delta \leftarrow 1.9(1-\gamma)/L_0$
$\boldsymbol{x}^1 \leftarrow \boldsymbol{x}^0$
for $k = 1, 2, \ldots, N_{\text{iter}}$ **do**
  $\boldsymbol{x}^{k+1} \leftarrow \mathbf{prox}_{\delta R}\left(\boldsymbol{x}^k - \delta\boldsymbol{\nabla}\mathrm{NLL}(\boldsymbol{x}^k;\boldsymbol{y}) + \gamma(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})\right)$
end for

We note that Algorithm 1 is an instance of the iPiano [22], although we have made some slight changes, where we chose this method due to its analysis with non-convex objective functions as we have in (3.9). It has been shown in [22] that iPiano will converge to a stationary point. However, due to its non-convexity, there is no guarantee that this will be a global minima. The parameter δ is the step size, which scales with a factor of L_0,

which we define as

$$L_0 = \left\| \boldsymbol{\Phi}^T \left[\left(\sum_{j=1}^{N_\xi} \alpha_1^2(\xi_j) \boldsymbol{b}(\xi_j) \right) \odot \boldsymbol{\Phi} \mathbf{1} \right] \right\|_\infty , \tag{3.13}$$

where $\| \cdot \|_\infty$ is the maximum norm of the vector. L_0 represents the maximum in the diagonal of the Hessian of (3.11) at the point $\mathbf{0}$, and may be considered as a crude estimate of the global Lipschitz constant, L. We note that this step size is likely to be very conservative in practice. The parameter γ sets the 'inertia' of the method, where we used $\gamma = 0.8$ as it resulted in the fastest performance in [22]. Finally, $\mathbf{prox}_{\delta R}$ is the proximity operator defined as

$$\mathbf{prox}_{\delta R}(\boldsymbol{z}) = \underset{\boldsymbol{\rho} \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{z} - \boldsymbol{\rho}\|_2^2 + \delta \lambda R(\boldsymbol{\rho}). \tag{3.14}$$

In our experimental section, we use the TV [98] as a regularisation function R, in which case (3.14) may be evaluated as in [104].

To illustrate the convergence properties of our method, and investigate its robustness to more aggressive step sizes, we applied Algorithm 1 to the chest data detailed in Section 3.4.2, and plotted the evolution of the NLL through iterates in Figure 3.3, for different step size multiplication factors. It follows that using δ is indeed very pessimistic,

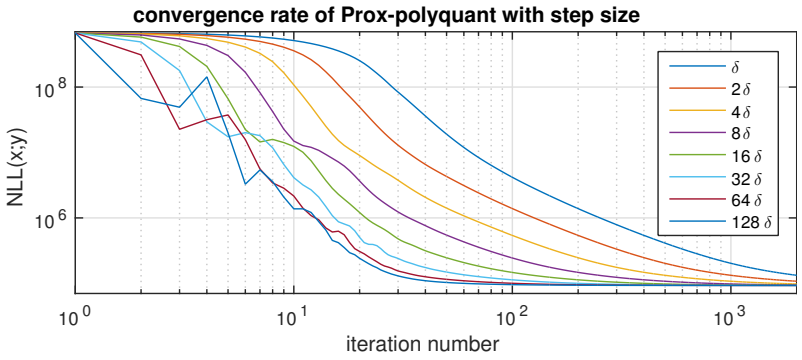


Figure 3.3: Convergence analysis for various step size multiplication factors without regularisation applied to the *chest data* detailed in Section 3.4.2. Each NLL here has a constant value added to it to aid visualisation.

and one observes convergence in this case even with 128δ, where convergence is achieved around 100 iterations. For factors larger than 128, we have found the method fails to converge, and it can be seen than the objective does not decrease monotonically for more aggressive step sizes. This analysis suggests that the local Lipschitz constant of the gradient through iterations is normally significantly less than the global maximum, or indeed the approximation with L_0. Adaptive step size schemes could exploit this fact,

such as the backtracking line-search in [22], but one should be aware that evaluating the cost function in (3.12) will be more than half the cost of calculating a new gradient, so each backtrack evaluation will increase the iteration cost significantly.

A remaining issue is the selection of N_{iter} in practice. One option is to set N_{iter} to the maximum allowable running time of the method, which will depend on the hardware and the imaging workflow. Another, is to terminate the algorithm once the residual, defined as the norm of the gradient, drops to some fraction of its initial size [102]. In this thesis however, we will be terminating the method at a point where the quality metric decreases by some amount smaller than the reported precision.

3.4 Evaluation from Simulation

3.4.1 Polyquant Model Evaluation

To investigate the accuracy of our proposed attenuation model, and its comparison to the other parameterisations given in Section 2.2.2, we calculated the predicted linear attenuation coefficient for a number of representative tissues from ICRP 89 [3]: adipose tissue, muscle bone, spongy tissue (upper femoral spongiosa), and hard bone.

For the DECT model in (2.6), we optimised the three scalar parameters K_1, K_2 and n for best fit to all materials in the ICRP 89—shown in Figure 3.1. We also fitted our piecewise linear Polyquant model in Section 3.2 for both relative electron density and monoenergetic attenuation at 60 keV to the same data. Similarly, we evaluated the accuracy of the IMPACT parameterisation in [11] to monoenergetic attenuation, again using the same fitting data from ICRP 89. These are plotted along with water and bone attenuation models in Figure 3.4 and the residual norms are tabulated in Table 3.1.

An observation that can be made from the plots in Figure 3.4 and results in Table 3.1 is that both IMPACT and our Polyquant models are very accurate over all materials. Fitting from monoenergetic attenuation as in IMPACT or 'Polyquant-tissue-μ_{mono}' is understandably more accurate than ρ_e, due to its closer similarity to other monoenergetic attenuations, and one will inevitably lose this accuracy if one later calibrates to ρ_e or ρ using the trend in Figure 2.7. We note that using our energy dependent fitting strategy in the case of monoenergetic attenuation consistently outperforms IMPACT, and this difference is an

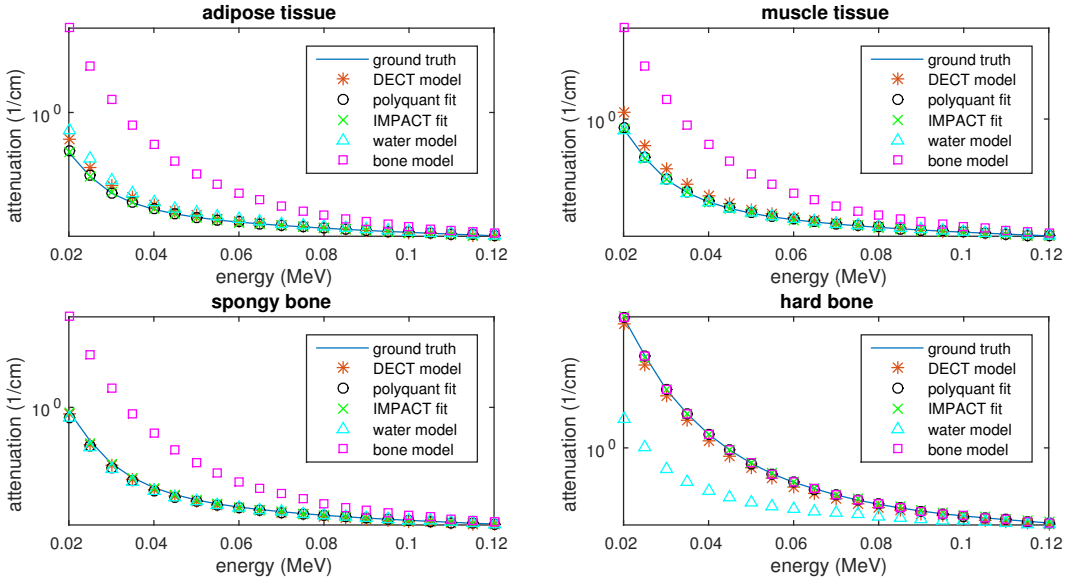


Figure 3.4: Model fitting results for tissue materials: adipose tissue, muscle tissue, spongy bone (upper femoral spongiosa), and hard (cortical) bone.

Table 3.1: Residual norm for various models fitted to various tissue materials.

Scheme	adipose tissue	muscle tissue	spongy bone	hard bone
DECT	0.15	0.28	0.047	0.97
Water	0.27	0.015	0.12	6.9
Bone	3.7	3.8	3.7	**0.0**
IMPACT	0.029	0.011	0.038	0.15
Polyquant-tissue-ρ_e	0.011	0.018	0.11	0.031
Polyquant-tissue-μ_{mono}	**0.010**	**0.0016**	**0.037**	0.0080

order of magnitude for the case of hard bone.

Another feature of the numerical results in Table 3.1 is that explicitly using the photoelectric–Compton relation in (2.6) as in DECT is less accurate than fitting it to biological materials as in IMPACT. This highlights that the physical photoelectric–Compton model is not very accurate over a wide range of material types, but also that one should not expect the fitted parameters from IMPACT to necessarily be physically meaningful in terms of ρ_e and Z_{eff}.

Finally, the water model is reasonably accurate for the soft tissues and spongy bone and not hard bone, and the bone model conversely so, which suggests the component-wise model in (2.9) is sensible. However, even if one selected the best cases from bone or water, as is the essence of [45–47], then the error would still be considerably higher than either

IMPACT or our proposed approach.

To demonstrate that these fitted parameters are not universal, we also ran the same models with the same fitted parameters on two synthetic plastic materials: polymethyl methacrylate (PMMA) and polytetrafluoroethylene (PTFE). These are interesting materials since they have similar electron densities to muscle and bone respectively but significantly different attenuation. In order to extend these models to include metallic implants, we also looked at how well they may model the attenuation of solid Titanium, where both IMPACT and Polyquant included a second 'knee' point to incorporate its attenuation also.

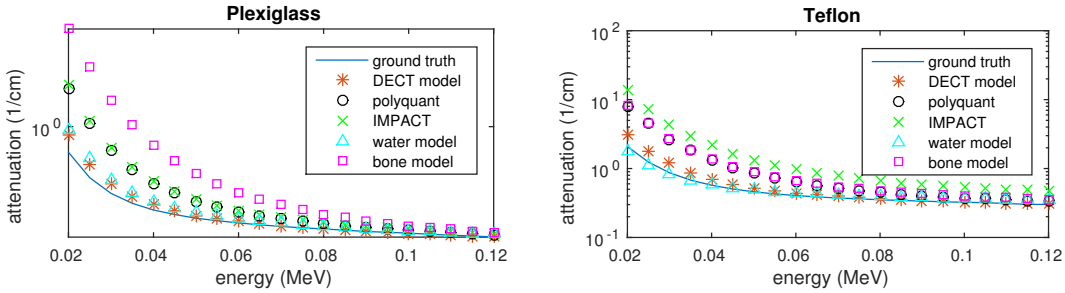


Figure 3.5: Model fitting results for synthetic plastic materials: PMMA and PTFE.

Table 3.2: Residual norm for various models fitted to PMMA and PTFE.

Scheme	Plexiglass	Teflon	Titanium
DECT	0.24	1.3	79
Water	0.35	0.40	82
Bone	4.6	7.5	66
IMPACT	1.5	14	1.9
Polyquant-tissue-ρ_e	1.3	7.2	0.032
Polyquant-tissue-μ_{mono}	1.5	14	**0.0076**
Polyquant-plastic-ρ_e	**0.020**	**0.0028**	84

From the plots in Figure 3.4 and results in Table 3.1, it is apparent that our proposed model is not universal across material types. Indeed, a water model is the best performing upon the specific plastic case, and the DECT is the best universal model, though this is also to be expected given the data in Figure 2.8 since it uses a two-dimensional parameterisation.

For the metallic implants, we note how the DECT, water and bone models are very inaccurate. IMPACT also shows significant errors, despite fitting the photoelectric–

Compton model directly to Titanium, which provides further evidence that the energy dependence model is not universal. On the other hand, by imposing no physical model, our Polyquant approaches are able to capture the attenuating profile of the implant material very closely.

The implications from the model experiment is that although our model is able to fit very closely to both tissue materials and metallic implants, once these parameters are fitted, they are inaccurate in synthetic plastic materials. An explanation for this mismatch is due to the fact that the correlation between electron density and attenuation coefficient is significantly different with tissues and synthetic materials [30]. With this, although parameterising the mass attenuation coefficient with water and bone gives more consistent predictions across material types, if one wishes to map into electron density for radiation therapy applications, then one must also take care to fit to the appropriate class of materials.

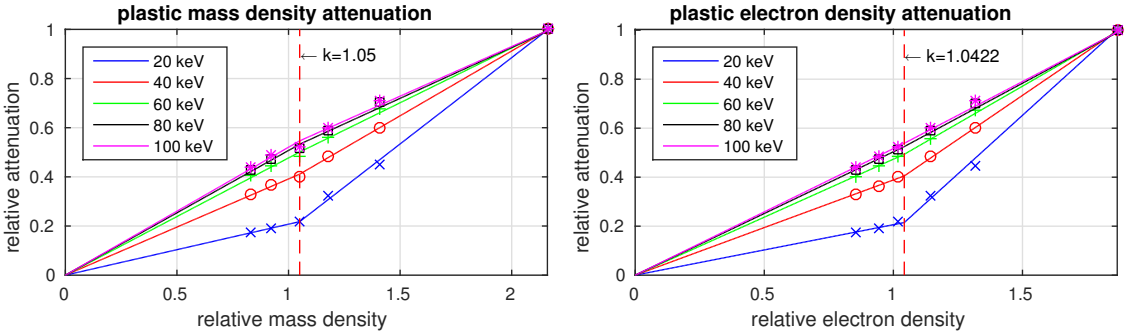
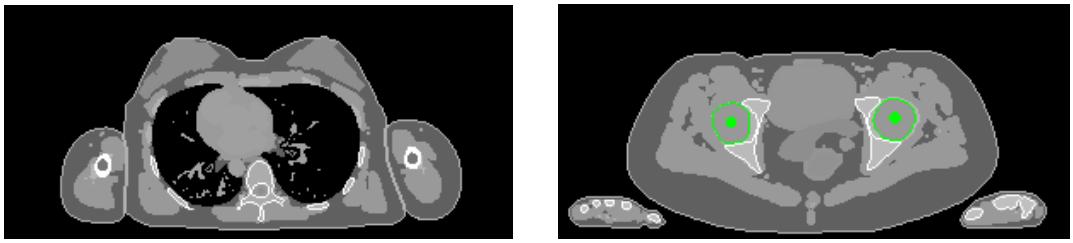


Figure 3.6: Relation between mass and electron density for a collection of synthetic materials. As with Figure 3.1, the attenuation has been normalised to the maximum attenuating material—Teflon in this case—to allow visualisation on a single graph

We finally note that our model is still able to account for plastic materials, and the 'Polyquant-plastic-ρ_e' in Table 3.2 is a result of fitting to a family of polymers—we used PTFE, polyoxymethylene (POM), PMMA, polystyrene, low-density polyethylene (LDPE) and polymethylpentene (PMP)—and the resulting accuracy is notably very high. We also show the relationship to attenuation for these materials in the graphs in Figure 3.6, which confirms that plastics interestingly follow a similar piecewise trend of their own, though this clearly does not extend to Titanium. If one wishes to quantify the attenuation of a mixture of both synthetic and biological tissues, then it seems that no single energy parameterisation would be consistent. One may opt for DECT measurements, but at the cost of requiring two diverse spectral sources.

3.4.2 Low Dose Numerical Reconstruction Test

The data we used for our numerical reconstruction test were derived from the Adult Reference Computational Phantom [41], which is a segmented image of defined density and chemical composition to represent real tissues. To investigate the ability to image metal implants, we inserted a pair of prosthetic hip joints (marked in green) with a solid titanium pin and shell. The slices through the chest and pelvis—which are the images we selected for testing—are shown in Figure 3.7. The resolution is 299×137.



(a) Chest test image (b) Pelvis test image

Figure 3.7: Experimental data used: (a) is the oracle chest electron density; and (b) is the oracle pelvis electron density; both have a with display window of [0.8,1.2]

The system geometry used for testing was a flat detector fan-beam CT, which was modelled in the Monte Carlo package Gate [26] as outlined Section 2.7. We included a focused lead strip collimator to mitigate scatter, a 0.6 mm thick CsI scintillator, a shaped source to put more photons into the centre of the specimen, and generated spectra using SpekCalc [143]: 6 mm aluminium filtration at 120 kVp. We simulated a total of 3×10^9 photons into 512 detection elements over 360 projections in 1° increments, which represented a low dose acquisition. Although the spectrum used to generate the measurements was sampled at 1 keV increments, for reconstruction we used $N_\xi = 21$, which was deemed a sufficient number in [11], and corresponded to our sampling of mass attenuation coefficients.

The implementation of all methods was in MATLAB, where we used separable footprint operators [70] from the Michigan Image Reconstruction Toolbox [153] for $\mathbf{\Phi}$ and $\mathbf{\Phi}^T$.

Reconstruction Methods Under Test

Most of the methods we tested are iterative reconstructions, with some degree of polyenergetic modelling. For fairness, we used TV regularisation for each [98], and selected the regularisation parameter giving the highest performance in each case — curves showing

the influence of these parameters are shown in Figure 3.9. All iterative methods were run for 2000 iterations, which resulted in empirical convergence in each case, after which the figure of merit did not changed by more than the reported precision.

Details of the methods under test are:

- *FBP*: Filtered backprojection (FBP) is a popular reconstruction algorithm, that 'analytically' approximates the inverse of linearised CT projections [55]. In our implementation: we take the logarithm of raw measurements; correct for beam hardening with a water only polynomial fitting [45]; apply a ramp filter with Hann windowing, with cut-off frequency optimised to give maximum performance; and finally apply a single backprojection operation $\boldsymbol{\Phi}^T$. To map to density from attenuation coefficient, we use the calibration curve in Figure 2.7 and detailed in [30].

  We included FBP as a crude baseline and to indicate the level of noise in our system, but we expect it to perform significantly worse than other competitive approaches under test due to the low dose.

- *PWLS*: Penalised weighted-least-squares (PWLS) approximates the CT model as linear by taking the logarithm of the raw measurements [60], calibrating them to correct for beam-hardening artefacts for the polyenergetic source [45], and includes a statistical weighting to approximate the Poisson noise in (2.13). The objective function is then

$$\boldsymbol{\mu}_{\text{mono}} = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} (\boldsymbol{\Phi}\boldsymbol{\mu} - \boldsymbol{l})^T \boldsymbol{W} (\boldsymbol{\Phi}\boldsymbol{\mu} - \boldsymbol{l}) + \lambda R(\boldsymbol{\mu}), \tag{3.15}$$

  where $\boldsymbol{W}$ is a diagonal statistical weighting matrix with entries $W_{ii} = (y_i - s_i)^2/y_i$, and $\boldsymbol{l}$ is the collection of linearised monoenergetic projections [60]. Converting $\boldsymbol{\mu}_{\text{mono}}$ to electron density is then done through a nonlinear calibration according to [30], as with the FBP. It should be noted that we are not actively modelling the metal implant in this case.

- *Poly-SIR*: Polyenergetic statistical iterative reconstruction (Poly-SIR) is the segmented water–bone model from (2.9) [46]. As this requires prior knowledge on material classes, we give it the oracle segmentation of the hard bone structures, and it treats everything else as water. In the pelvis case, we also pass the oracle segmentation of the metal implants, with corresponding mass attenuation

coefficients. Since the physical model gives the mass density, we convert to relative electron density where appropriate using the curve shown in Figure 2.8b.

- *IMPACT*: We implement the IMPACT model of [11] as a special case of our generalised fitting in (3.6). This allows us to use the same algorithm as presented in Section 3.3 for its minimisation. We use a three component piecewise-linear fitting from monoenergetic attenuation at 60 keV to energy dependent attenuation, which accounts for metal implants as well as biological tissues. We then use the same post-processing calibration technique as with FBP and PWLS to convert to either mass or relative electron densities.

- *Polyquant*: In our proposed model, we use the piece-wise linear fitting in Section 3.2 and reconstruction strategy in Section 3.3. We use Algorithm 1, but with 10δ which from Figure 3.3 is still rather conservative, but exhibited a monotonic decrease in the objective function. For the pelvis case, we extend the model with a second 'knee' point and linear section to include the attenuation of solid titanium according to the generalisation in (3.2). We used separate fittings to electron and mass density shown in top and middle graphs in Figure 3.1, to demonstrate the ability to reconstruct directly into either quantity, where each mapping was a least-squares fitting to the materials in ICRP 89 [3]. The resulting parameters for the relative electron density fitting are given in Appendix A.

Quality Assessment Metric

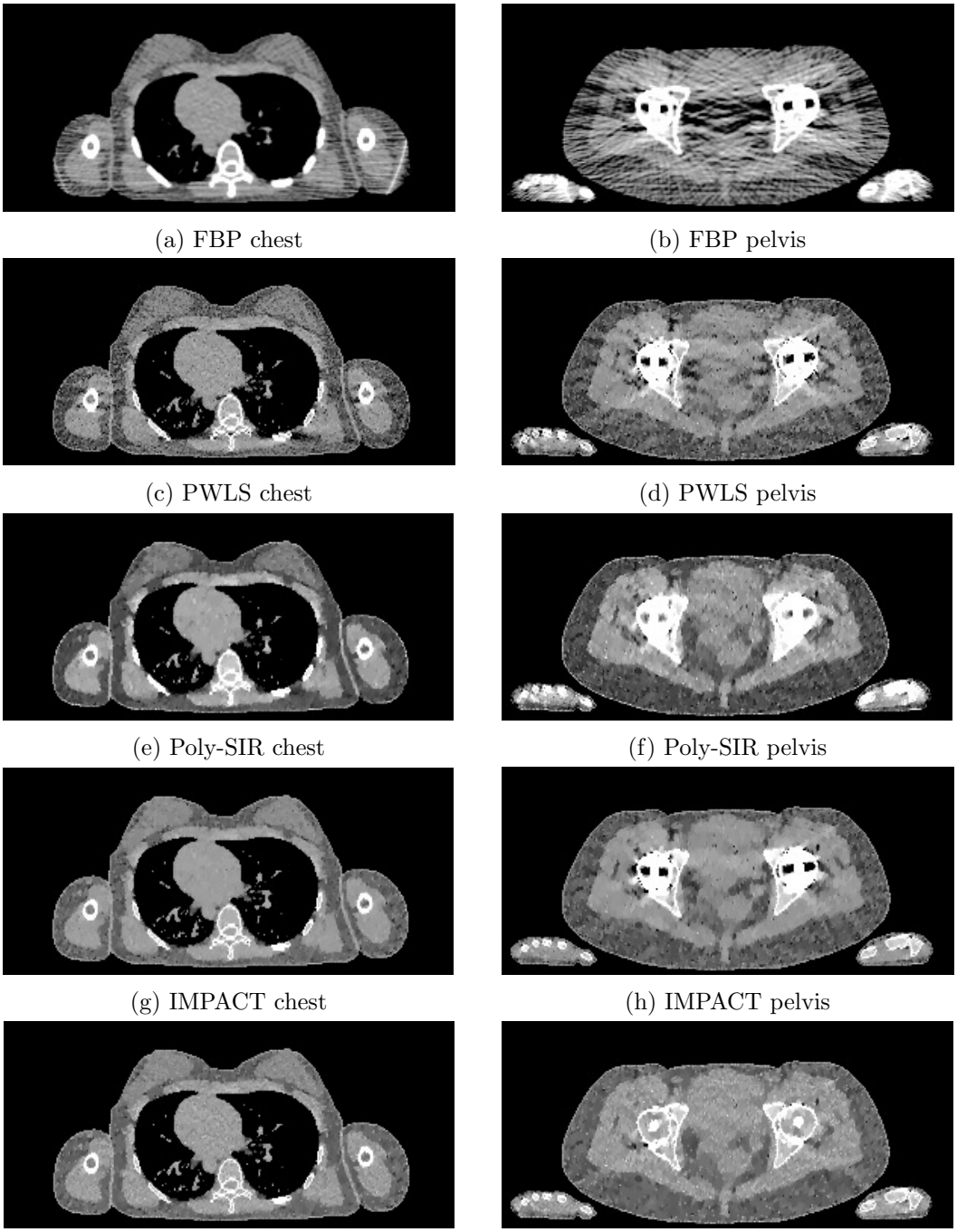
The metric for quantifying the accuracy of the various methods under test is the root-mean-squared-error (RMSE) in relative electron and mass density, calculated as

$$\sqrt{\frac{1}{N_{\text{vox}}}\sum_{i=1}^{N_{\text{vox}}}(\hat{\rho}_i - \rho_i)^2}, \tag{3.16}$$

where $\hat{\boldsymbol{\rho}} \in \mathcal{R}^{N_{\text{vox}}}$ is the estimated mass or electron density of a tested method, and $\boldsymbol{\rho}$ is the ground-truth. As an error, the lower the score represents a higher quantitative accuracy.

Reconstruction Results

The results from our reconstruction test are illustrated in Figure 3.8 and Table 3.3.



(a) FBP chest (b) FBP pelvis

(c) PWLS chest (d) PWLS pelvis

(e) Poly-SIR chest (f) Poly-SIR pelvis

(g) IMPACT chest (h) IMPACT pelvis

(i) Polyquant chest (j) Polyquant pelvis

Figure 3.8: Results from electron density reconstruction test for chest and pelvis fan-beam data with display window [0.8,1.2] to aid visualisation of soft tissue and reconstruction artefacts. The associated oracle images are those in Figure 3.7.

In terms of both electron density and mass density accuracy, our proposed method is the best performing method under test. The lack in estimation accuracy in using the water–bone model with Poly-SIR is likely to be due to discrepancies in adipose and spongy bone tissues from water, which were illustrated in Figure 3.4. We note that Poly-SIR is second

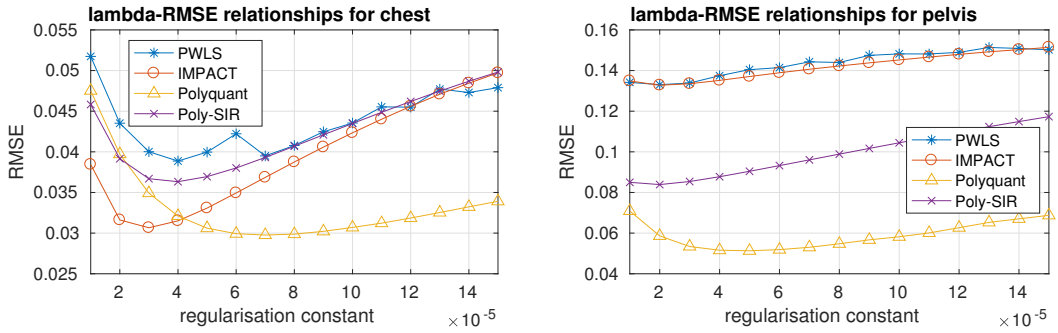


Figure 3.9: Curves of electron density RMSE against regularisation parameter for iterative methods under test in chest and pelvis data sets. The minima from each curve is selected in each case given the reconstructions shown in Figure 3.8 and quantitatively assessed in Table 3.3

Table 3.3: Quantitative results: relative electron density (ρ_e) and mass density (ρ) RMSE

	chest data		pelvis data	
Scheme	ρ_e RMSE	ρ RMSE	ρ_e RMSE	ρ RMSE
FBP	0.136	0.140	0.201	0.226
PWLS	0.0389	0.0426	0.133	0.166
Poly-SIR	0.0363	0.0385	0.0839	0.101
IMPACT	0.0307	0.0326	0.133	0.159
Polyquant	**0.0298**	**0.0316**	**0.0513**	**0.0746**

best performing in the pelvis case, but it was provided with oracle information on the implant and hard bone, which would be difficult to segment in practice from a preliminary reconstruction such as FBP due to its high noise. IMPACT is understandably very closely performing to our method in the chest case, as these performed similarly in the model test, but the advantage of fitting at each energy is clear in the case of the metal implant, where the IMPACT performance is similar to that of PWLS.

The relationship between regularisation strength and quantitative performance of the iterative methods are shown in Figure 3.9. In both cases, this demonstrates that the Polyquant not only reaches the best performance, especially in the pelvis case, but its numerical accuracy is reasonably robust to the setting of this parameter. The difference in the location of the minima between methods may be accounted for by the different scales in image parameter — for example IMPACT uses the monoenergetic attenuation at 60 keV, whereas Polyquant uses the physical density, and these have a relative difference in intensities. Another implication from these curves is that although the Polyquant chest

image in Figure 3.8i exhibits a higher level of intratissue variation than IMPACT, this will be mitigated by increasing the regularisation parameter, which can increase by over 40% whilst maintaining the best numerical performance.

Another interesting feature of the results is the difference between electron and mass density scores. Although it may seem counter-intuitive that the Poly-SIR for example would have a higher error before calibration than after, this is due to the RMSE being absolute and not relative. Since from Figures 2.8b and 3.1 the mass densities of materials are on average higher than relative electron density, this should account for the difference.

3.5 Accelerating the Prox-polyquant Algorithm

The algorithm given in Algorithm 1 performs reconstruction through gradient descent optimisation of the Polyquant objective function. Although this appears robust in empirical convergence even under aggressive step-size multiplication factors, as shown in Figure 3.3, each iteration is potentially expensive to compute, as it requires several full projection and back-projection calculations. For the numerical experiment in Section 3.4.2 with single slice fan-beam CT, this cost was reasonable, but for CBCT data it is likely to be considerably higher in both memory and computation. For this reason, we next investigate and analyse algorithmic acceleration strategies that will be employed in the following experimental section.

As was outlined in Section 2.5.4, one can realise acceleration through calculating the gradient on a subset of the measurements. One method therefore, is to apply Prox-polyquant as given in Algorithm 2, with each iteration only a acting on a subset of the full NLL. Adopting the same notation as the prototypical subset gradient method in (2.46), this NLL can be expressed as

$$\mathrm{NLL}(\boldsymbol{x}^k; \boldsymbol{y}, \mathcal{S}^k) = \sum_{i \in \mathcal{S}^k} \sum_{j=1}^{N_\xi} \psi_i(\boldsymbol{x}, \xi_j) + s_i - y_i \log \left(\sum_{j=1}^{N_\xi} \psi_i(\boldsymbol{x}, \xi_j) + s_i \right),$$

given some subset $\mathcal{S}^k \subset [1, \ldots, N_{\mathrm{ray}}]$. As in (2.47), the projection indices array is defined as

$$\boldsymbol{\iota}^k = [l^k : N_{\mathrm{split}} : (N_{\mathrm{proj}} - N_{\mathrm{split}} + l^k)], \tag{3.17}$$

where $\boldsymbol{\iota}$ is a subset index array that selects a subset as

$$\hat{\boldsymbol{y}}_j = [\boldsymbol{y}\{\boldsymbol{\iota}^k\}]_j \text{ for } j = 1, \ldots, \frac{N_{\text{proj}}}{N_{\text{split}}}, \tag{3.18}$$

where $\hat{\boldsymbol{y}}$ is the subset of the full set of measurements $\boldsymbol{y}$. From this, a subset variant of the Prox-polyquant is given in Algorithm 2.

Algorithm 2 Fast-Polyquant 1: heavy ball momentum

Initialization: Subset selection scheme for $\mathcal{S}^k$, regularisation constant λ, Lipschitz approximation L_0, number of iterations N_{iter}, number of subset divisions N_{split}, inertia factor $\gamma \in [0, 1)$, and starting point $\boldsymbol{x}^0$.
$\boldsymbol{x}^0 \leftarrow \mathbf{1}$
$\delta \leftarrow 1.9(1-\gamma)/L_0$
$\boldsymbol{x}^1 \leftarrow \boldsymbol{x}^0$
for $k = 1, 2, \ldots$ **do**
  $\boldsymbol{x}^k \leftarrow \mathbf{prox}_{\delta R}\left(\boldsymbol{x}^k - \delta\boldsymbol{\nabla}\text{NLL}(\boldsymbol{x}^k; \boldsymbol{y}, \mathcal{S}^k) + \gamma(\boldsymbol{x}^k - \boldsymbol{x}^{k-1})\right)$
end for

In other ordered subset realisations as in [23, 114, 119], the authors employ diagonal preconditioning on the objective function, which is the precomputed diagonal of the Hessian at point $\boldsymbol{x} = \mathbf{0}$ in those cases. With the non-convex NLL of CT, these precomputed preconditioning do not guarantee a monotonic algorithm, although Erdogan and Fessler [114] demonstrate that it performs well empirically. We can use this same concept also, by defining a factor

$$\boldsymbol{p} = \boldsymbol{\Phi}^T \left[\left(\sum_{j=1}^{N_\xi} \alpha_1^2(\xi_j) \boldsymbol{b}(\xi_j) \right) \odot \boldsymbol{\Phi}\mathbf{1} \right], \tag{3.19}$$

It can be seen that the parameter L_0 in (3.13) is the maximum of this vector. To apply this, the scalar step-size factor δ is replaced by

$$\boldsymbol{\delta} = 1.9(1-\gamma) \oslash \boldsymbol{p}, \tag{3.20}$$

which is now a vector that individually weights each element of the gradient. This changes the gradient step calculation by

$$\delta\boldsymbol{\nabla}\text{NLL}(\boldsymbol{x}^k; \boldsymbol{y}, \mathcal{S}^k) \rightarrow \boldsymbol{\delta} \odot \boldsymbol{\nabla}\text{NLL}(\boldsymbol{x}^k; \boldsymbol{y}, \mathcal{S}^k). \tag{3.21}$$

By including preconditioning on forward–backward splitting, the proximal term should also be modified to account for this [111]. We will analyse if this diagonal preconditioning offers any gain in performance, though only with non-negativity constraints, whereby the

proximal operator need not be changed.

We also wish to compare the 'heavy ball' inspired acceleration [22,117] against the Nesterov momentum [118] used in FISTA [10] shown in (2.53). This combination of Nesterov momentum and subset gradient descent was analysed in [23] and demonstrated in [119] for CBCT reconstruction. This combination applied to our objective function results in Algorithm 3.

Algorithm 3 Fast-Polyquant 2: FISTA acceleration

Initialization: Subset selection scheme for $\mathcal{S}^k$, regularisation constant λ, Lipschitz approximation L_0, number of iterations N_{iter}, number of subset divisions N_{split}, and starting point $\boldsymbol{x}^1$.
$\boldsymbol{x}^1 \leftarrow \mathbf{1}$
$\boldsymbol{z}^0 \leftarrow \boldsymbol{x}^1$
$\delta \leftarrow 1.9N_s/L_0$
for $k = 1, 2, \ldots$ **do**
$\boldsymbol{z}^k \leftarrow \mathbf{prox}_{\delta R}\left(\boldsymbol{x}^k - \delta\boldsymbol{\nabla}\text{NLL}(\boldsymbol{x}^k; \boldsymbol{y}, \mathcal{S}^k)\right)$
$t^{k+1} \leftarrow \frac{1+\sqrt{1+4(t^k)^2}}{2}$
$\boldsymbol{\rho}_e^{k+1} \leftarrow \boldsymbol{z}^k + \left(\frac{t^k-1}{t^{k+1}}\right)(\boldsymbol{z}^k - \boldsymbol{z}^{k-1})$
end for

Finally, an element that proved to have a significant impact on the stability of these types of methods is the ordering of the subsets [23]. The best performing method that we observed was the 'bit-reversal' (BR) ordering of [7]. To illustrate this, Figure 3.10 shows a BR selection for $N_{\text{split}} = 32$. It can be seen that each subsequent index is a large distance from the last.

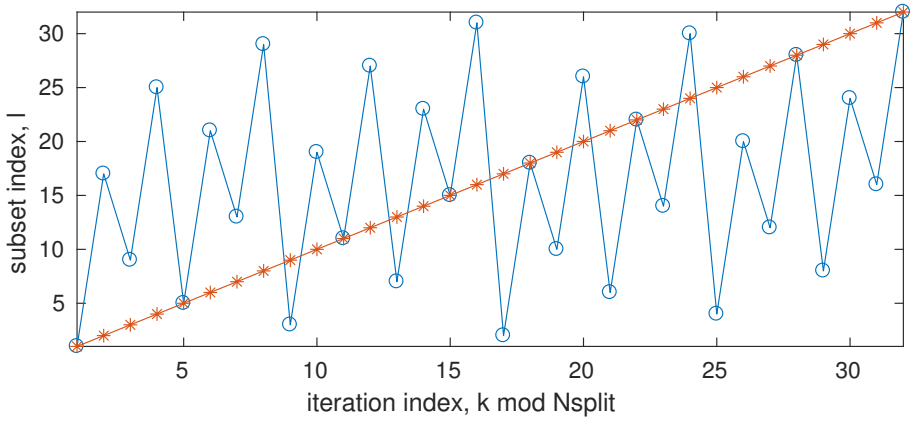


Figure 3.10: Demonstration of 'bit-reversal' ordering scheme applied to $N_{\text{split}} = 32$: the red line marked with * shows sequential ordering; the blue line marked with shows the BR ordering from [7].

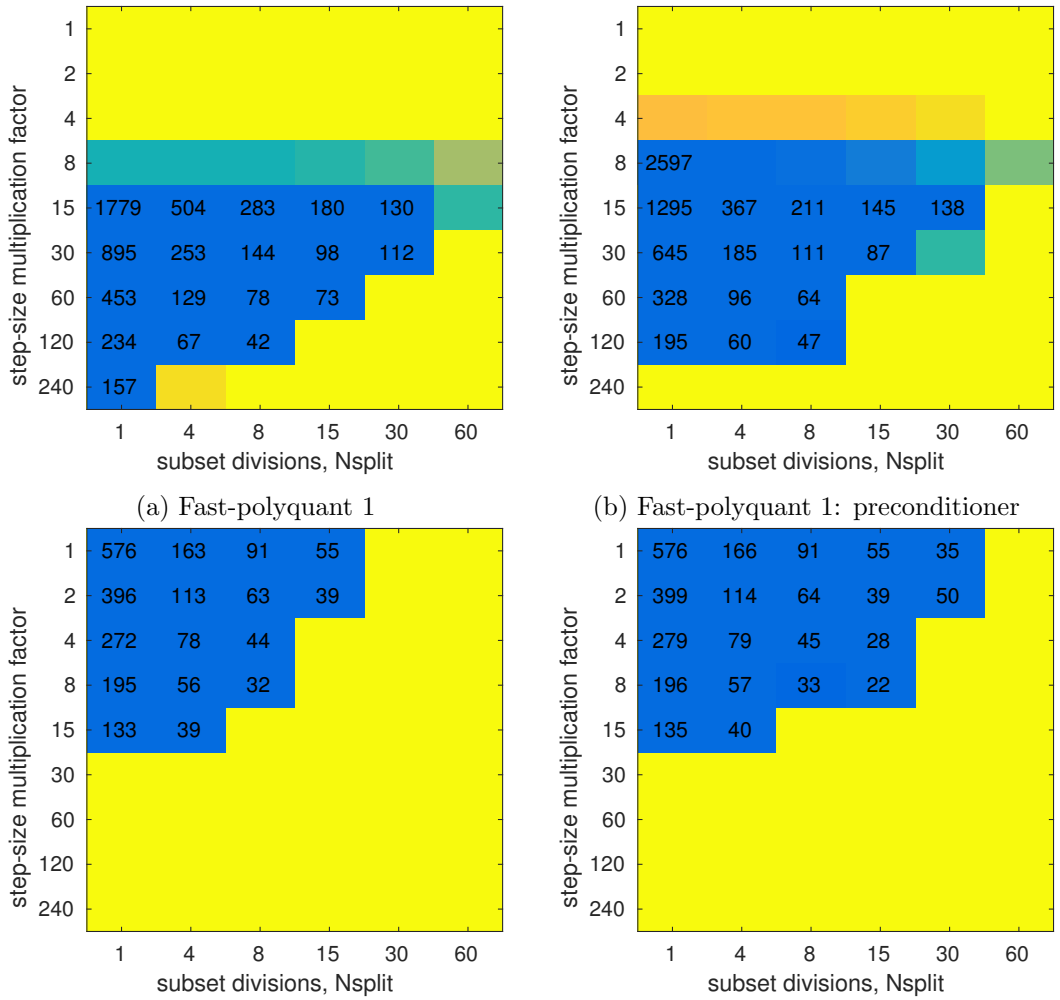
3.5.1 Analysis and Validation

To analyse the relative performance of the various elements of algorithm acceleration, we devised a benchmark to evaluate the speed of convergence in each case. As with the convergence curves plotted in Figure 3.3, this will compare speed under different step-size multiplication factors, but also different subset divisions, and with no regularisation. This is distinct from the analysis in [23], where only different subset sizes were compared.

The data used for this investigation was the numerical chest data as used in Section 3.4.2. We characterised a method as being successful, if it could achieve an object function value less than 9.5×10^4 — selected from the final values achieved in Figure 3.3 — within 2000 iterations, N_{iter}. The results from this test are illustrated in Figure 3.11. We evaluated: Fast-polyquant 1 (F-P 1) as given in Algorithm 2; Fast-polyquant 1 with diagonal preconditioning from (3.19); and Fast-polyquant 2 (F-P 2) as given in Algorithm 3 with both sequential and BR subset selection. Methods that successfully achieved an objective value less than 9.5×10^4 are shown in blue, along with the time in seconds they took to reach this, with methods that failed to converge due to either instability or slowness in yellow. We highlight that the original Prox-polyquant (P-P) in Algorithm 1 is represented as the first column in Figure 3.11a.

For analysis purposes, we calculated the full NLL at each iteration in every case. Since calculation of this full objective function requires $2N_f - 1$ applications of $\mathbf{\Phi}$, this corresponds to a huge cost, which is more than the expense of calculating a full gradient. In practice, the objective function would not be evaluated in subset techniques, as this would negate any speed increase. Therefore, in the calculation of the execution times in Figure 3.11, we have subtracted this objective calculation time.

There are many interesting trends that can be found from Figure 3.11. Firstly, there is a consistent relationship between stability, step-size factor, and subset divisions. In can be seen that as one increases the step-size factor, that the maximum successful subset splitting factor decreases. One marked difference between the two acceleration variants, is the F-P 1 from Algorithm 2 appears active at considerably higher step-sizes than the FISTA variant of Algorithm 3. This could well be related to the step-size scaling factor of $(1 - \gamma)$ in the former case, where we used $\gamma = 0.8$. Although this was required to prove convergence in [22], these results suggest that this may not be necessary in practice. A



(a) Fast-polyquant 1

(b) Fast-polyquant 1: preconditioner

(c) Fast-polyquant 2: sequential

(d) Fast-polyquant 2: bit-reversal

Figure 3.11: Grids showing the time various methods take to reach an NLL value of 9.5×10^4 for the numerical chest data. The number indicates the executing time in seconds minus the objective calculation time for successful instances. The colour map has windowing of $[9.4 \times 10^4, 1 \times 10^5]$ (from blue to yellow) after a maximum of 2000 iterations.

similar effect was also observed in Figure 3.3, where convergence was observed for very high step-size scaling factors. The reason the lowest step-sizes are represented in yellow as a failure in Figure 3.3a and Figure 3.3b, was simply that they were too slow to reach the desired objective value within the 2000 iterations.

By comparing the results between Figure 3.3a and Figure 3.3b, this suggests that using diagonal preconditioning is not effective for our objective function. Although its inclusion does offer significant acceleration factors for the full gradient with $N_{\text{split}} = 1$, this begins to diminish as the subset divisions are increased. For the most aggressive acceleration factors, the preconditioning even becomes detrimental.

The effect of using BR ordering can be seen by comparing Figure 3.3c with Figure 3.3d. Up to $N_{\text{split}} = 8$ there is no practical difference between the ordering strategies, with each giving identical performance. What BR appears to offer over sequential ordering however, is stability with much more aggressive subset divisions.

Between the acceleration strategies, although both have similar properties, the FISTA momentum of Prox-polyquant 2 does appear to exhibit higher stability, and it is likely easier to select an appropriate step-size multiplication factor for this in practice — even using 1δ offers a better performance than any of the tested cases of Prox-polyquant 1.

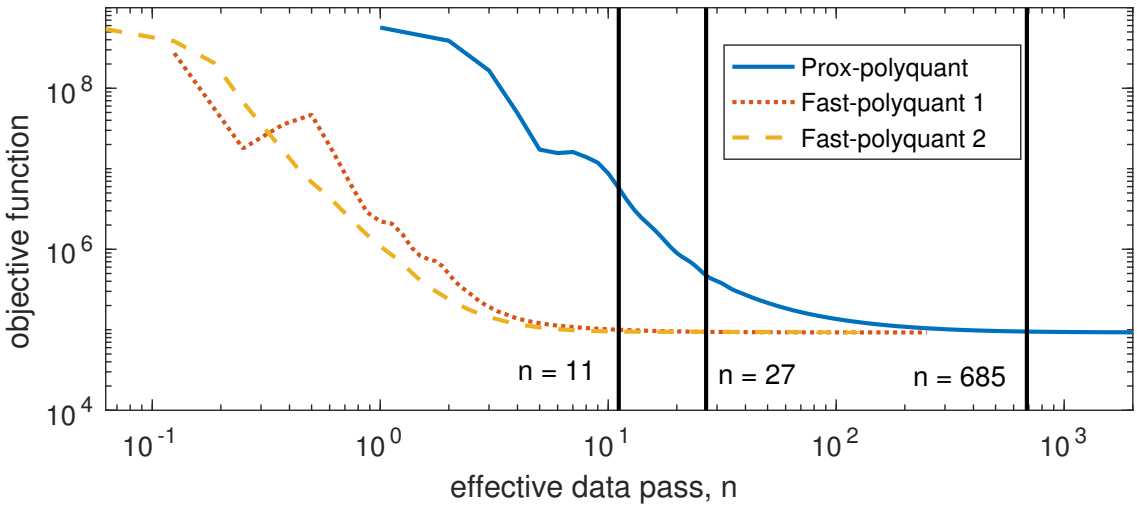
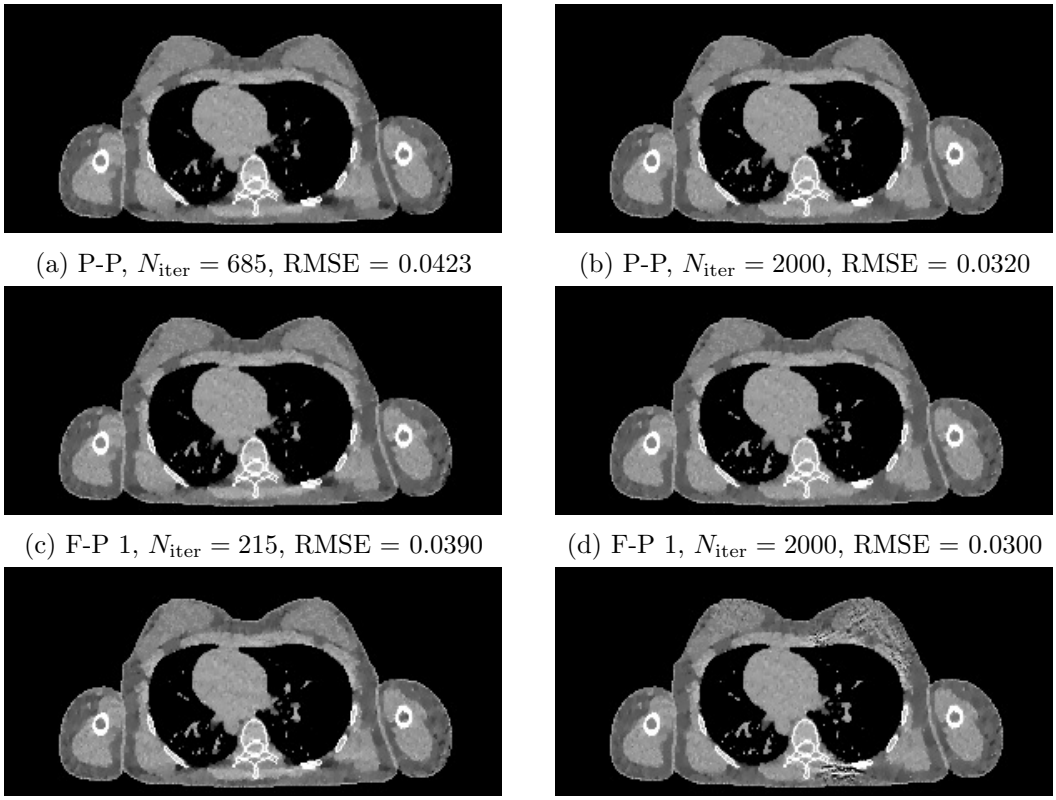


Figure 3.12: Curves showing objective function against effective data passes (epochs) for various algorithm types on the numerical chest data. All are run until their termination for NLL value of 9.5×10^4.

To see if either of the acceleration strategies, F-P 1 or F-P 2, began to diverge after many iterations, we ran select examples for $N_{\text{iter}} = 2000$. We selected: F-P 1 with 8 subset divisions and step-size factor of 120 that reached NLL $= 9.5 \times 10^4$ in 42 seconds at $N_{\text{iter}} = 215$ in Figure 3.11a; F-P 2 with 15 subset divisions and step-size factor of 8 that reached termination in 22 seconds at $N_{\text{iter}} = 178$ in Figure 3.11c; we also used a representative full Prox-polyquant implementation with step-size multiplication factor of 30, giving a monotonically decrease in objective function, and termination with 685 iterations in 895 seconds in Figure 3.11a. The curves for each case for $N_{\text{iter}} = 2000$ are shown in Figure 3.12, with the data pass where each method was terminated in Figure 3.11 also shown.

From Figure 3.12, neither of the methods appear to diverge in terms of objective function, even after many iterations with this case of no regularisation. As a more practical

experiment, we also analysed the relative reconstruction performance, with the addition of TV regularisation. Here we ran each of the three selected methods from Figure 3.12, both for the number of iterations that achieved termination with NLL $= 9.4 \times 10^4$, and for 2000 iterations. The resulting reconstructions are shown in Figure 3.13.



(a) P-P, $N_{\text{iter}} = 685$, RMSE $= 0.0423$ (b) P-P, $N_{\text{iter}} = 2000$, RMSE $= 0.0320$

(c) F-P 1, $N_{\text{iter}} = 215$, RMSE $= 0.0390$ (d) F-P 1, $N_{\text{iter}} = 2000$, RMSE $= 0.0300$

(e) F-P 2, $N_{\text{iter}} = 178$, RMSE $= 0.0347$ (f) F-P 2, $N_{\text{iter}} = 2000$, RMSE $= 0.0340$

Figure 3.13: Illustrative results from fast algorithm analysis. Each is shown with display window [0.8,1.2].

The regularised reconstruction results in Figure 3.13 highlight some interesting properties. Firstly, at the points of NLL termination in Figure 3.11, which is equivalent to $N_{\text{epoch}} =$ 685, 27, 11 respectively, all three methods exhibit similar levels of artefacts and structure preservation. At this point F-P 2 also interestingly has significantly the lowest RMSE. At $N_{\text{iter}} = 2000$ however, a different pattern emerges. Whilst Prox-polyquant and F-P 1 achieve similar levels of image structure and RMSE, F-P 2 shows considerable artefacts in the lower and upper right sections. This may either be an effect of the nonconvexity of the objective function, where this method may be stuck in a local minima, or an effect of the regularisation function. In any case, we are yet to confirm that F-P 1 is generally less prone to these artefacts, and this would have to be the subject of further study.

There are a couple of conclusions that can be drawn from this section. Firstly, subset

gradient descent appears to offer huge gains in computation time over full gradient. Diagonal preconditioning appears to be detrimental in the subset setting, whilst BR subset ordering offers a significant enhancement in stability. The Fast-polyquant 2 method, using FISTA like acceleration, is extremely fast, offers good preservation in image structure with only 11 data passes, and is relatively stable with aggressive acceleration. However, it has the potentially to produce undesirable image artefacts if left to run for many iterations. Although Fast-polyquant 1 has a more modest acceleration and may pose a difficulty in selecting a suitable step-size multiplication factor, it was more stable in our regularised reconstruction test. We suggest that if executing time is a strong constraint, then F-P 2 would be a good option, but F-P 1 may be a more robust choice given more computational resource.

3.6 Physical Phantom Validation

To validate our method, we also tested its ability to perform quantitative reconstruction from real X-ray CBCT measurements. For this, we acquired a scan with a Varian TrueBeam On-Board Imager of two anthropomorphic phantoms: an SK200 Sectional Chest Phantom manufactured by the Phantom Laboratory, incorporated; and a CIRS STEEV head phantom.

The SK200 consists of a real human skeleton cast in a synthetic resin and a textured synthetic lung material. Since the resin material has a measured mass density, we will assess the accuracy of reconstruction into this. Secondly, the CIRS STEEV phantom similarly consists of synthetic materials, but which are manufactured to match the relative electron densities of human tissue. Therefore, we are able to quantify the relative electron density accuracy in this case.

3.6.1 SK200 Chest Phantom

SK200 Data Processing

Our CBCT acquisition consists of 888 projections at a 125 kVp tube potential and 15 mA current for 20 mSec on each. The system also included a Titanium beam hardening filter, a bow-tie filter, a scatter collimation grid, and a Cesium Iodine scintillator. The

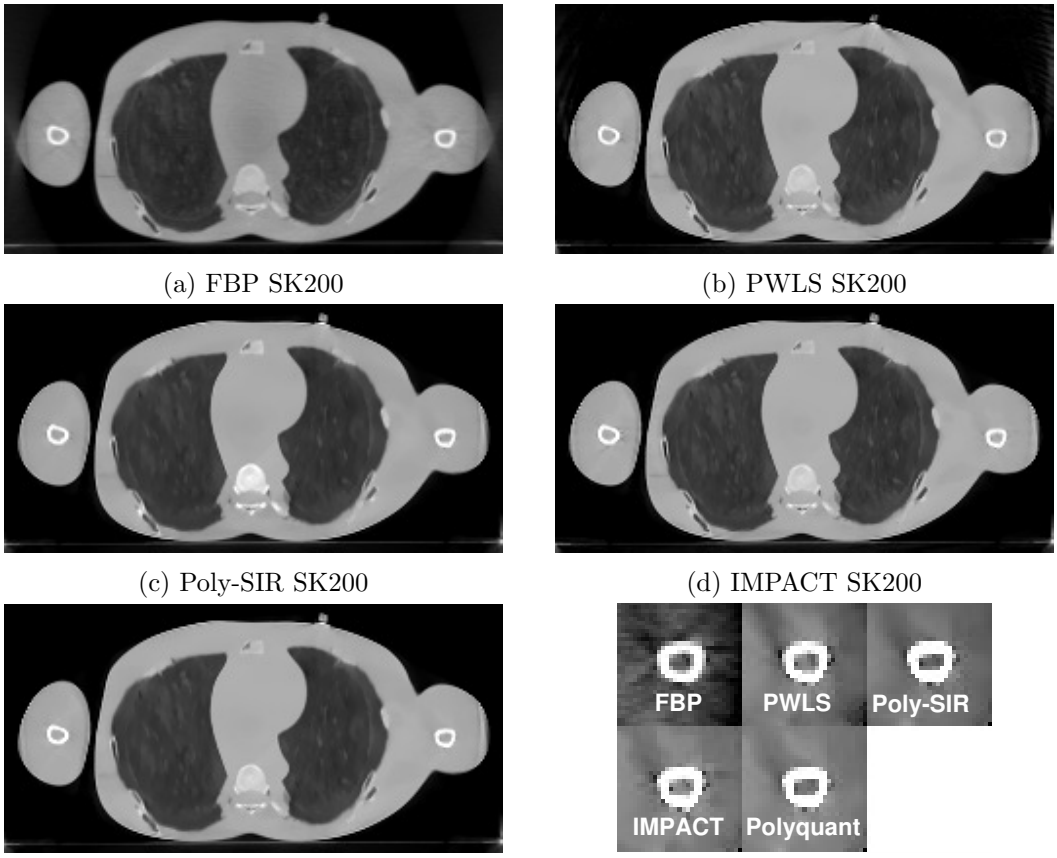
raw measurements were all pre-corrected for detector responses, and the effect of bow-tie shifting with gantry rotation, with default TrueBeam$^{\text{TM}}$ corrections to give $\boldsymbol{y}$ in (2.13). The scatter estimate $\boldsymbol{s}$ in (2.13) was also taken from the Varian system's default scatter correction. For testing the FBP—realised with the Feldkamp–Davis–Kress method [9]—and PWLS, we took the linearised projection vector $\boldsymbol{l}$ in (3.15) after the system's beam hardening correction and logarithm.

In the case of the fully polyenergetic reconstruction methods Poly-SIR, IMPACT and Polyquant, we require knowledge of the X-ray spectrum $b_i(\xi)$ in (2.13). Due to the variable thickness of Aluminium in the bow-tie filter, this will be spatially varying, and we calculated it analytically from the appropriate spectrum in the system's calibration parameters, and the spectral response of the various metal filters and scintillator. As in our other experimental sections, we discretised the spectrum into 21 energies, and used the same parameters as were fitted in Section 3.4.1 for the ICRP biological tissues.

SK200 Reconstruction

For reconstruction, we mapped into a resolution of $256 \times 256 \times 96$, and used each method as detailed in Section 3.4.2. In each case, we used offset detector weighting as in [57]—although this should not be necessary for iterative methods, we noted it mitigated artefacts in the centre of these reconstructions. We ran each iterative method for 200 iterations. With lack of ground-truth data to inform its optimisation, we heuristically used $0.2\lambda_{\text{chest}}$, where λ_{chest} were the same TV regularisation parameters from the digital chest experiment, and gave good empirical performance on the CBCT data. Finally, for the bone segmentation required for Poly-SIR, we applied a threshold of $\rho > 1.5\,\text{g/cm}^3$ to the FBP reconstruction, which empirically isolated the hard structures in the image, and assigned these voxels as bone.

Reconstructions of the 57^{th} slice from the CBCT data are shown in Figure 3.14. We note that each image looks rather similar structurally, due to a high number of measurements, and larger relative dose in comparison to our numerical experiment in Section 3.4.2. To illustrate relative differences, we have also shown the isolated region around a bone in Figure 3.14f. Here, the Polyquant result shows the lowest distortion around the bone. There are still visible artefacts in the soft tissues in this case, and this may be improved further with more accurate scatter estimation or spectrum modelling. On the other hand,



(a) FBP SK200 (b) PWLS SK200

(c) Poly-SIR SK200 (d) IMPACT SK200

(e) Polyquant SK200 (f) Bone

Figure 3.14: Results from mass density reconstruction from real CBCT data, where each is shown with display window [0,1.5].

the FBP does exhibit a drop in intensity in Figure 3.14f that may be attributed to the insufficient data sampling with CBCT [18]

Figure 3.15: (a) Mask corresponding to muscle-like resin in the 57^{th} slice, from segmentation; (b) Expanded views of bone structure within arm, with display window [0.8,1.2] to highlight artefacts.

To evaluate the quantitative accuracy of each reconstruction, we isolated voxels in the 57^{th} slice corresponding to the muscle-like resin. This is shown in Figure 3.15, where there are 8024 voxels here. The RMSE and standard deviation of these voxels to the manufacturer's

quoted mass density of $1.01\,\mathrm{g/cm^3}$ are tabulated in Table 3.4.

Table 3.4: Quantitative CBCT results: root-mean-squared-error (RMSE) between tissue voxels and measured mass density of $1.01\,\mathrm{g/cm^3}$, and corresponding standard deviation (std).

Scheme	FBP	PWLS	Poly-SIR	IMPACT	Polyquant
RMSE	0.127	0.0396	0.0403	0.0269	**0.0244**
std	0.0460	0.0357	0.0265	0.0250	**0.0239**

From Table 3.4, we note that our proposed model is nearly 10% more accurate than the next best method under test. Additionally, the variation of our method is also the lowest, which should be indicative of less soft-tissue distortion from insufficient modelling of bone attenuation.

3.6.2 STEEV Head Phantom

The CIRS STEEV phantom consists of synthetic resins to mimic the attenuating properties of human tissues, allowing quantitative assessment of relative electron density accuracy. There was also a metal structure in the centre of the phantom, consisting of the plug section from a PTW PinPoint ionisation chamber, allowing us to investigate the mitigation of metal induced artefacts.

STEEV Data Processing

Our CBCT acquisition consists of 499 projections at a 100 kVp tube potential and 20 mA current for 15 mSec on each, which were the default settings for a head acquisition. Compared to the numerical test, the relative X-ray flux was roughly 2.5× higher, which coupled with the smaller specimen volume and larger number of projections, implies that this test was at a significantly higher dose.

As in our other experimental sections, we discretised the spectrum into 21 energies, and used the same parameters as were fitted in Section 3.4.1 for the ICRP biological tissues, but supplemented by the mass attenuation of the metal implant according to information provided by the manufacturer.

STEEV Reconstruction

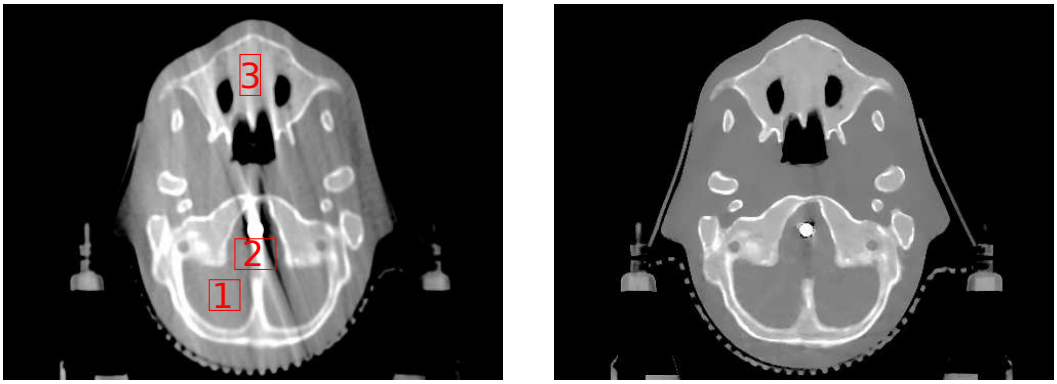
For reconstruction, we mapped into a resolution of $512 \times 512 \times 144$, and used each method as detailed in Section 3.4.2. We ran each iterative method for 500 iterations. For the regularisation parameter, we heuristically used $0.5\lambda_{\text{pelvis}}$, where λ_{pelvis} were the same TV regularisation parameters from the digital pelvis experiment, and gave good empirical performance on the CBCT data. Finally, for the bone and metal segmentations required for Poly-SIR, we obtained these through applying thresholds on the FBP and PWLS separately. To illustrate the critical role of this step, we have shown both images in Figures 3.16c and 3.16d.

Reconstructions of the 83$^{\text{rd}}$ slice from the CBCT data are shown in Figure 3.16, along with a region from the 91$^{\text{st}}$ slice containing a higher mass of metal in Figure 3.17. Although the FBP appears to suffer strongly from the presence of the metal structure, most of the iterative methods mitigate its effect considerably, with the PWLS showing a more pronounce dark region in the soft tissue. The Poly-SIR based on a segmentation from the FBP does suffer from its streaking as shown in Figure 3.16d. Although a better performance is achieved through segmenting from the PWLS as in Figure 3.16d, this will have a considerably higher total computational cost; the Poly-SIR region in Figure 3.17c is based on this PWLS initialisation. The preservation of bone structure appears to be similar between all the iterative methods. The Polyquant method appears to have the smoothest regions around the metal implant, whilst maintaining the same level of structure in the bone and objects surrounding the phantom.

To evaluate the quantitative accuracy of each reconstruction, we isolated regions of interest (ROI) in the 83$^{\text{rd}}$ slice located in soft tissue and spongy bone — these are shown in Figure 3.16a. The RMSE of these regions are calculated relative to the electron density of the soft tissue and bone equivalent resins, and shown in Table 3.5.

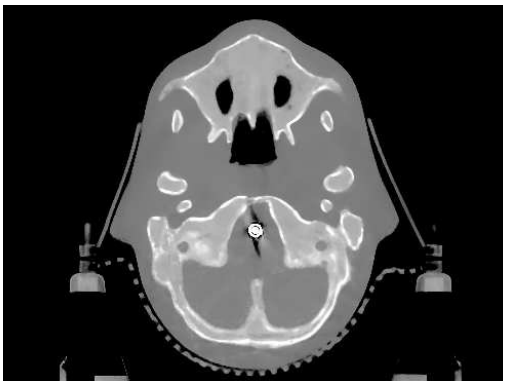
Table 3.5: Quantitative CBCT results: RMSE of relative electron density in regions shown in Figure 3.16a.

RMSE	FBP	PWLS	Poly-SIR	IMPACT	Polyquant
ROI 1	0.0583	0.0170	0.0200	0.0154	**0.0100**
ROI 2	0.231	0.0235	0.0256	0.0199	**0.0166**
ROI 3	0.0342	0.0376	0.0333	0.0178	**0.0128**

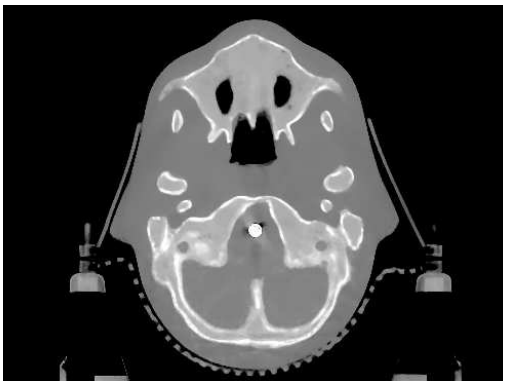


(a) FBP CBCT

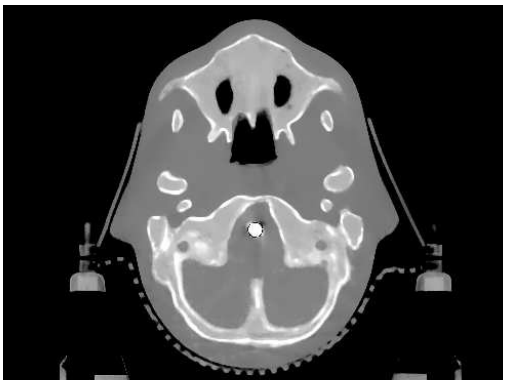
(b) PWLS CBCT



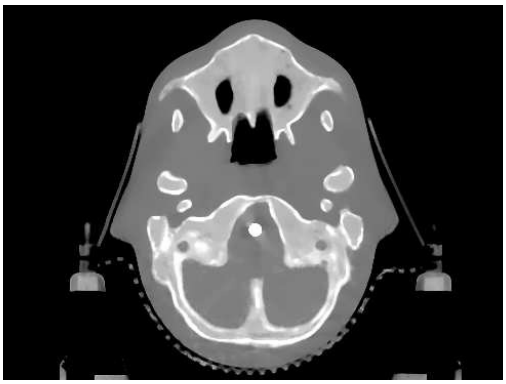
(c) Poly-SIR CBCT 1



(d) Poly-SIR CBCT 2



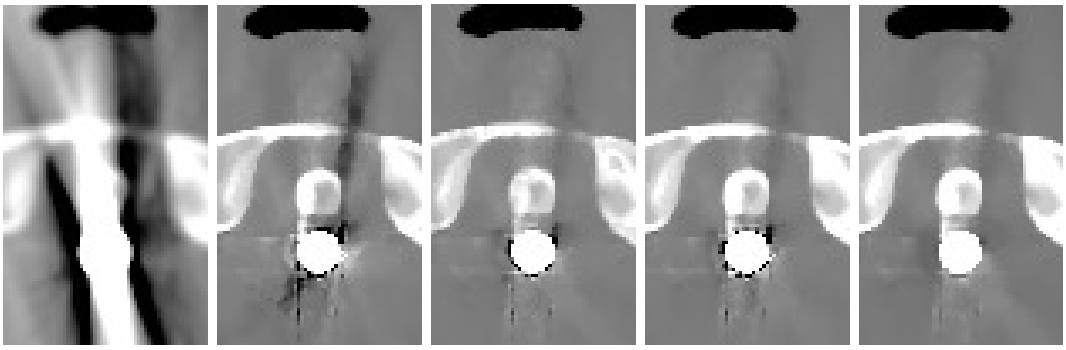
(e) IMPACT CBCT



(f) Polyquant CBCT

Figure 3.16: Results from electron density reconstruction from real CBCT data showing slice 83, where each is shown with display window [0.7,1.4]: (c) is Poly-SIR given a bone and metal segmentation derived from the FBP; (d) uses a segmentation derived from the PWLS.

From Table 3.5, we note that our proposed model is the most accurate method under test by at least 16% over other approaches in all regions.



(a) FBP (b) PWLS (c) Poly-SIR (d) IMPACT (e) Polyquant

Figure 3.17: Metal insert visualisation of slice 91 from CBCT reconstructions with display window [0.7,1.4]. The Poly-SIR is the version based upon segmentation from the PWLS as that in Figure 3.16c.

3.7 Discussion

Aspects of our method we have not evaluated in this study are its robustness, and the practicality of its computational implementation, though these are both worth discussing. In the first case, we have noted that due to the discontinuous gradient from the piecewise linear fitting, there is no theoretical guarantees for convergence. Two approaches we have adopted in implementations are using a smooth function for $f(x)$ in (3.5) such as a generalised logistic, and connecting the two linear fits with a quadratic function for some interval around the 'knee'. However both options increase the computational cost of the gradient term in (3.11) considerably, and we have found neither give any empirical advantage in convergence or accuracy over just using the non-smooth version. Since it is common in CT reconstruction to use empirically well performing methods that have no convergence guarantees such as pre-computed curvatures in separable paraboloid surrogates [86] or ordered subsets [114], we believe this is reasonable. We also highlight, that from Figure 3.3, our method does empirically appear to converge even for very aggressive step size multiplication factors.

Another potential robustness issue with any polyenergetic model are partial volume effects, where a discretised voxel contains different classes of material. Due to its linearity, given that the two materials belong to the same fit interval, such as lung and fat or muscle and bone as in Figure 3.1, then our model would correctly estimate the attenuation from these materials. If a voxel contains materials from different intervals however, such as fat and muscle or metal and bone, then our model will overestimate the density, according to the

trends in Figure 3.1. Although we have not evaluated the degree of this effect, we note it will be in common with other approaches [11, 46, 47], and may be mitigated by increasing the resolution of reconstruction.

The performance of iterative methods under a low dose acquisition is an important consideration for reducing the amount of ionising radiation delivered to the patient. By performing our numerical test at a low dose, we demonstrated the method is relatively robust in this setting. Between the iterative methods tested, PWLS is expected to suffer the most from a low dose, due to its approximation to the noise model, and linearisation of the projections that become unstable for very low photon fluxes [60], and has been shown to perform worse at low doses than 'pre-log' methods such as Polyquant, Poly-SIR and IMPACT [154]. Between these three methods, whether there exists a relative change of interplay between partial volume effect photon flux and accuracy is yet to be determined, although Polyquant is likely to benefit from its superior attenuation modelling.

In terms of computational cost, our method is comparable to other iterative methods. Given that the bottleneck is in calculating the forward- and back-projection operators $\mathbf{\Phi}$ and $\mathbf{\Phi}^T$, then we note that each gradient step in (3.11) may be implemented with 3 forward and 2 backward evaluations, for biological materials, which represents a 2.5× larger cost than PWLS — this is the same cost as our implementation of IMPACT. Comparatively, Poly-SIR has a 2× larger cost than PWLS, which given the consistent accuracy advantage of our model, is unlikely to be worth this slight speed advantage. By employing accelerated algorithms, described in Algorithm 2 and Algorithm 3, we have demonstrated dramatic speed increases of up to 30×. If these methods were implemented on parallel hardware, such as GPUs, then we believe they would be feasible for on-line reconstruction, whilst the patient is lying on the couch.

3.8 Conclusions

We have introduced a general quantitative attenuation model, which allows direct inference of mass or electron density from raw CT measurements with a single polyenergetic source. Not only have we demonstrated this allows more accurate modelling than explicit physical models such as water–bone or photoelectric–Compton, but have shown how it may be exploited in a flexible reconstruction algorithm that allows accurate quantitative medical

imaging, even with metal implants and real CBCT data. As with other single source methods, we have highlighted its inconsistency between synthetic and biological tissues, but this may not be of relevance for medical imaging, in which opting for the more general DECT model is significantly less accurate over materials of interest. Since our method has a similar computational cost to other iterative approaches, but offers markedly higher accuracy, it offers both a practical and beneficial approach to CT imaging.

Chapter 4

Quantitative Reconstruction with Scatter Model Fusion

4.1 Introduction

X-ray scatter is a large source of errors in CT, where scattered X-rays corrupt the line of sight attenuation models used during reconstruction. In CBCT, due to its large field of view, the magnitude of these interfering X-rays is commonly of the same order of magnitude as the signal of interest and can even be considerably higher [137]. With this, artefacts and inaccuracies in the reconstruction are inevitable unless it can be correctly compensated. In Chapter 3, we introduced a polyenergetic quantitative reconstruction technique allowing the direct inference of electron density — known as Polyquant. This method can incorporate a prior estimate of the additive scatter, but this does not exploit any of the polyenergetic information or indeed the electron density available during iterations.

The problem then remains of how one should generate this estimate of scatter, which is an active area of research with many possible approaches [128]. A fundamental reason why estimating scatter is difficult is that unlike modelling attenuation, requiring a single path from source to detector for each measurement, one must consider every possible path a photon can take through various numbers of scattering events to get the full picture. This will therefore not only be dependent on the attenuating materials and intensity along a single pencil beam, but it will be dependent on the full projection fluence and the complete structure of the specimen. Not only would calculating all these paths be computationally exhaustive, but one does not even have knowledge of the structure of the specimen prior to reconstruction, since this is the task of the reconstruction itself. This therefore presents a dependency problem, where one requires the image to estimate scatter, but cannot form the image accurately because of scatter contamination.

To overcome this limitation, a compelling approach is to integrate the scatter estimation and reconstruction processes, so that they are performed simultaneously. This notion is exploited to some degree by efficient Monte Carlo scatter estimation methods such as [131–134], where one requires a preliminary reconstruction to estimate the scatter, which can then be iteratively refined. Another approach in [135] used an analytic single scatter model to generate scatter estimates during reconstruction. A drawback of either Monte Carlo or analytic scatter models is their computational cost, which would make the already high cost of iterative reconstruction huge.

Another family of scatter estimation approaches that are purely measurement based are convolutional methods, as were described in Section 2.6.2, known as scatter kernel superposition (SKS) [8] or beam scatter kernels (BSK) [128]. Here the scatter is estimated as a convolution between a stationary point spread function, usually calculated or measured for a slab of material, and a forward scatter factor [136] derived from the small angle scatter magnitude hitting the line-of-sight detector element. SKS methods are typically very fast, due to their ability to exploit the fast Fourier transform (FFT) based calculation of convolution, but are naturally limited in their inability to account for inhomogeneities in the scatter paths. These methods may however be supplemented with heuristic perturbations to compensate for the most significant inhomogeneous effects such as fringing [155], varying thickness [8], effects at the edge of an object [8], scatter from the couch [156], and more general spatial inhomogeneities [157].

From a preliminary study comparing scatter estimation methods when used with statistical iterative reconstruction [158], we found that SKS approaches matched the performance of sub-sampled Monte Carlo estimates, whilst being significantly faster. For this reason, we focus on the integration of SKS models into the polyenergetic reconstruction algorithm, Polyquant, and see if their performance can be enhanced further from information available during the iterations.

In this work, we derive a polyenergetic SKS (PolySKS) model in terms of attenuation and electron density projections, include perturbations to account for the specimen position and inhomogeneities in scatter paths, and integrate this into the Polyquant framework. We denote this integrated reconstruction method as Polyquant–PolySKS. We then propose a fast algorithm for its implementation using accelerated ordered subsets as in [23], able to include sparsity exploiting regularisation such as total variation (TV). Finally,

we demonstrate the method on both numerical and real CBCT data, and analyse its performance in terms of relative electron density accuracy.

4.2 Methodology

4.2.1 Polyenergetic Scatter Modelling

Polyenergetic Forward Scatter Derivation

The derivation of the forward scatter factor for SKS for preprocessed estimation in Section 2.6.2 made the assumption that the incident beam was monoenergetic, and was in terms of just the linearised attenuation coefficient μ. However, one can derive this in a more accurate polyenergetic fashion, where the proportion of scattered X-ray at the point D in Figure 2.19 at a given energy ξ is

$$I_D(\xi) = I_U(\xi)K(\xi)\rho_e(U)\exp\left(-\int_U^D \boldsymbol{\mu}(\xi,\ell)\,d\ell\right). \tag{4.1}$$

There are a couple of differences here from (2.68). These are the explicit dependence on source energy ξ, and the dependence on the energy independent electron density ρ_e. The reason for this is that the attenuation through scattering events may be expressed in terms of the associated interactive cross sections as [39]

$$\mu_{\text{scatter}}(\xi) = \rho N_g(\sigma_{\text{incoh}}(\xi) + \sigma_{\text{coh}}(\xi)), \tag{4.2}$$

where $\sigma_{\text{incoh}}(\xi)$ is incoherent Compton scatter, and $\sigma_{\text{coh}}(\xi)$ are coherent scatter events. From the definition of relative electron density in (2.4) the proportionality to relative electron density can be seen. Furthermore, one can find the scalar constant $K(\xi)$ is related to the physics as

$$K(\xi) \propto \frac{\sigma_{\text{incoh}}(\xi) + \sigma_{\text{coh}}(\xi)}{\rho_{\text{water}} N_{g,\text{water}}}, \tag{4.3}$$

with the constant of proportionality being related to the size and position of the detector element. With this, having the scatter intensity proportional to μ as in (2.68) can be seen as an approximation and will implicitly include the non-scattering photoelectric interactions.

Similarly to the monoenergetic derivation in (2.69), the forward scatter factor can be calculated by integrating all the individual scatter factors from O to D, which can be expressed as

$$\begin{aligned} p(\xi) &= \int_O^D I_O(\xi) \exp\left(- \int_O^{U'} \boldsymbol{\mu}(\xi, \ell)\, d\ell \right) K(\xi) \rho_e(U') \exp\left(- \int_{U'}^D \boldsymbol{\mu}(\xi, \ell)\, d\ell \right) dU' \\ &= I_O(\xi) \exp\left(- \int_O^D \boldsymbol{\mu}(\xi, \ell)\, d\ell \right) K(\xi) \int_O^D \boldsymbol{\rho}_e(\ell)\, d\ell, \end{aligned} \tag{4.4}$$

where the combination of the two attenuation integrals can only be made since photons do not change energy with scattering events with no deflection [39, 159]. Since in practice one will have a finite detector size, there will be some small change in photon energy from Compton scatter, but the approximation in (4.4) is still reasonable since the angles will be small. In any case, the energy dependency of the scatter factor $K(\xi)$ is retained to allow any compensation for this, despite the zero angle scattering not having a dependency on energy [40].

As in (2.70), one can express this in a discretized setting for all measurements as

$$p_i(\xi) = K(\xi) b_i(\xi) \exp(-[\boldsymbol{\Phi}\boldsymbol{\mu}(\xi)]_i)[\boldsymbol{\Phi}\boldsymbol{\rho}_e]_i \text{ for } i = 1, ..., N_{\text{ray}}. \tag{4.5}$$

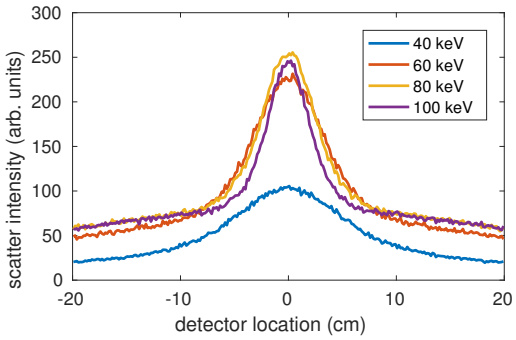
A convenient property of (4.5) when used within Polyquant reconstruction is that the energy dependent attenuation $\boldsymbol{\mu}(\xi)$ and the electron density $\boldsymbol{\rho}_e$ are readily available at each iteration without need for further computation. With this, given that scatter kernels are consistent within a given energy, then the same scatter model will be able to be used for any spectrum, given that is it known.

As with the conventional SKS derivation, (4.5) does ignore multiple scattering that rejoins the line–of–sight path, and we will analyse in the next section how well this assumption holds.

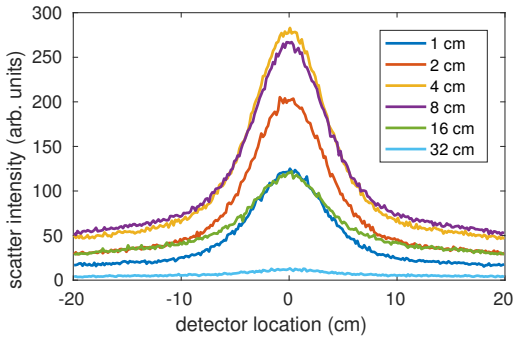
Polyenergetic Water Kernel Analysis

To build a convolutional model to utilise the polyenergetic scatter decomposition in (4.5), we must investigate the nature of kernels from different energy sources. For this, we used the Monte Carlo particle simulation toolbox Gate [26], to calculate scatter profiles

through slabs water of varying thickness but uniform cross sectional size of 80×80 cm. We positioned slab 50 cm from the planar detector and 100 cm from the point source. Some examples of these kernels are shown in Figure 4.1.



(a) 10 cm water slab kernels for different energies



(b) 60 keV scatter kernel for different slab thickness

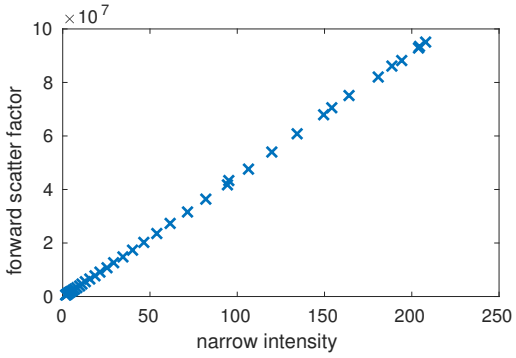
Figure 4.1: Examples of water slab scatter kernels from simulation plotted as detection intensity after detector response function and ADC (arb. units) against detector location (cm) along the u dimension as shown in Figure 2.19.

Initial observations from Figure 4.1 are that there is indeed a strong link between the shape of the scatter profile and energy of the photon, and it appears that the width of the kernel decreases with an increase in energy as in Figure 4.1a. As expected, there is also a strong link between thickness and kernel shape as shown in Figure 4.1b, though it appears the width is roughly preserved. This suggests that instead of using a multi thickness decomposition for different kernel shapes as in (2.73), one could instead decompose in terms of energy. This would fit with the polyenergetic forward model derived in (4.5).

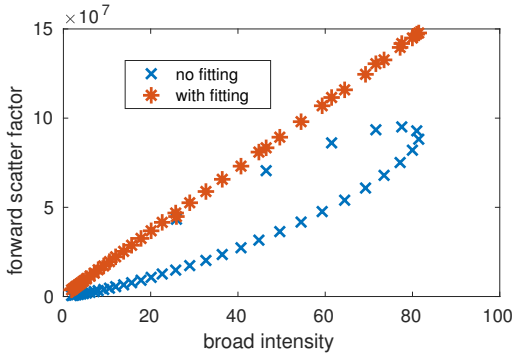
We have also plotted the relationship between the amplitude of the measured scatter kernels and the polyenergetic scatter factor with $K = 1$ as calculated in (4.5), which are shown in Figure 4.2. These are shown as scatter plots for an increasing thickness of material, where one would expect a perfect match to the model as a straight line through the origin.

It is clear from Figure 4.2a, that our analytic model fits extremely well to the narrow Gaussian, where one can perform fitting by simply finding an appropriate proportionality constant $K(\xi)$. On the other hand, it does not agree entirely with the broad Gaussian as shown in Figure 4.2b.

However, by employing the same adjustment parameters h_1, h_2 as in classical scatter factor calculation in (2.71) [136], we are able to obtain a very good linear fit also, which is shown



(a) Analytic scatter factor against narrow amplitude



(b) Analytic scatter factor against broad amplitude

Figure 4.2: Relationship between narrow analytic forward scatter factors in (4.5) and measured kernel amplitudes on water scatter data of 1–40 cm at 60 keV.

in Figure 4.2b. This then takes the form

$$p_{\mathcal{B},i}(\xi) = K(\xi) b_i(\xi) \exp(-[\mathbf{\Phi}\boldsymbol{\mu}(\xi)]_i)^{h_1} [\mathbf{\Phi}\boldsymbol{\rho}_e]_i^{h_2} \text{ for } i = 1, ..., N_{\text{ray}}. \tag{4.6}$$

Due to properties of exponentials, this may be rewritten as

$$p_{\mathcal{B},i}(\xi) = K(\xi) b_i(\xi) \exp(-h_1 [\mathbf{\Phi}\boldsymbol{\mu}(\xi)]_i + h_2 \log[\mathbf{\Phi}\boldsymbol{\rho}_e]_i) \text{ for } i = 1, ..., N_{\text{ray}}, \tag{4.7}$$

where the parameters become scaling factors, lending itself to faster computational implementation. The fact that our derived single scatter forward model fits so well to the narrow Gaussian amplitude may be indicative that these result from single scatter events, whilst the broad signal is attributed to multiple scatters. However, through analysing the scattering order from our simulations, we have found this not to hold entirely, with single scatter also contributing significantly to the broad peak.

With these and other observations, we note that double Gaussian kernels from various energies and thickness of slab have the following properties:

- The double Gaussian decomposition fits well for each energy and water thickness.
- The narrow field has a width that is dependent on the energy of the incident photons, and an amplitude that is directly proportional to the polyenergetic forward scatter factor in (4.5).
- The broad field has a width which is independent of photon energy, and is proportional to the polyenergetic forward scatter factor in (4.5) with appropriate

fitting parameters h_1, h_2.

- With the above factors taken into account, there appears no additional dependence on thickness, suggesting the polyenergetic decomposition takes account for the change in shape observed in [8].

These observations are key to the implementation details of our scatter model given in the next section.

Basic Polyenergetic Convolutional Model

Using the polyenergetic forward model in (4.5) and separate narrow and broad components, a basic form of double Gaussian polyenergetic scatter kernel accurate for a static slab may be written as

$$\boldsymbol{s}\{k\}(u,v) = \sum_{j=1}^{N_\xi} \boldsymbol{p}_{\mathcal{N}}\{k\}(u,v,\xi_j) * \boldsymbol{g}_{\mathcal{N}}(u,v,\xi_j) + \tag{4.8}$$

$$\boldsymbol{p}_{\mathcal{B}}\{k\}(u,v,\xi_j) * \boldsymbol{g}_{\mathcal{B}}(u,v,\xi_j). \tag{4.9}$$

Since we have found that the broad field is roughly independent of energy, this may be simplified to

$$\boldsymbol{s}_{\mathcal{B}}\{k\}(u,v) = \left(\sum_{j=1}^{N_\xi} \boldsymbol{p}_{\mathcal{B}}\{k\}(u,v,\xi_j) \right) * \boldsymbol{g}_{\mathcal{B}}(u,v), \tag{4.10}$$

which requires only a single convolution.

As with the pre calculated SKS, the convolution operations may be replaced by point-wise multiplication in the spatial frequency domain, whereby the calculation of scatter may be made as follows

$$\boldsymbol{s}\{k\}(u,v) = \mathcal{F}^{-1}\left[\boldsymbol{S}_{\mathcal{N}}\{k\}(\omega_u,\omega_v) + \boldsymbol{S}_{\mathcal{B}}\{k\}(\omega_u,\omega_v)\right] \text{ for } i = 1, ..., N_{\text{proj}}, \tag{4.11}$$

where (ω_u, ω_v) are the spatial frequencies along the dimensions of the detector plane (u, v), and

$$\boldsymbol{S}_{\mathcal{N}}\{k\}(\omega_u,\omega_v) = \sum_{j=1}^{N_\xi} \mathcal{F}\left[\boldsymbol{p}_{\mathcal{N}}\{k\}(u,v,\xi_j)\right] \odot \mathcal{F}\left[\boldsymbol{g}_{\mathcal{N}}(u,v,\xi_j)\right], \tag{4.12}$$

and

$$\boldsymbol{S}_{\mathcal{B}}\{k\}(\omega_u, \omega_v) = \mathcal{F}\left[\sum_{j=1}^{N_\xi} \boldsymbol{p}_{\mathcal{B}}\{k\}(u, v, \xi_j)\right] \odot \mathcal{F}\left[\boldsymbol{g}_{\mathcal{B}}(u, v)\right]. \tag{4.13}$$

We note that with this, the computational cost of a full scatter estimate involves N_{proj} IFFT operations and $N_{\text{proj}}(N_\xi + 1)$ FFT operations, since the Fourier transform of the Gaussian kernels need not be explicitly computed.

4.2.2 Accounting for Detector Distance

We already highlighted the fact that the forward scatter scalar constant $K(\xi)$ will in practice depend on the distance to the detector due to a finite detector size. On top of this, the width of the kernels will also vary as the distance to detector is varied, due to the magnification effect. In traditional preprocessing scatter calculation such as fASKS, these factors cannot be easily accounted for, since one does not know the position of the object. In our proposed integrated model however, this information is available from the current iterate, and may be exploited to further enhance the estimation accuracy.

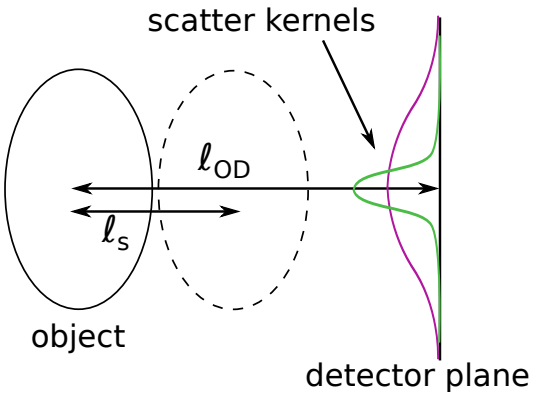


Figure 4.3: Scatter shift diagram illustrating magnification effect assuming the scattered photons originate from the centre of the specimen.

With reference to Figure 4.3, we can derive these correction factors based on the relative location of the object's centre of mass. This makes the assumption that the mean scatter originates from the centre of mass, which implies independence on the position of the source. As the object moves closer to the detector, we expect a decrease in Gaussian width, and an increase in amplitude due to the increase in area covered by the central

pixel. By introducing a magnification factor ζ, defined as

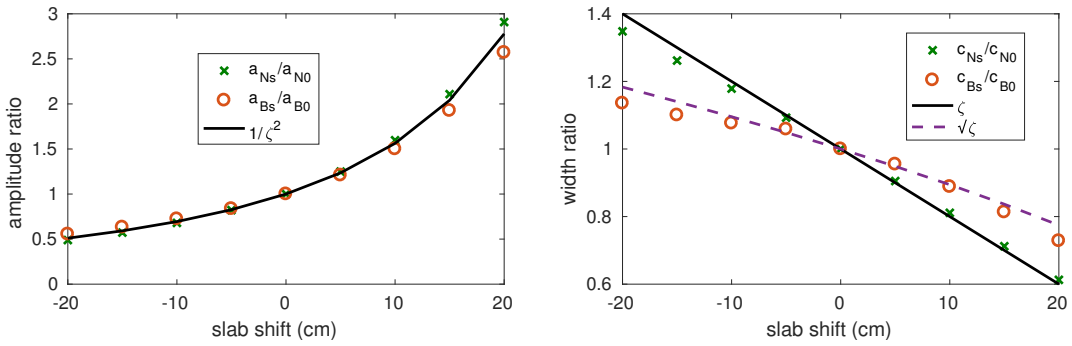
$$\zeta = \frac{\ell_{OD} - \ell_s}{\ell_{OD}}, \tag{4.14}$$

one would expect the width to change by a factor of ζ, and the amplitude by a factor of $1/\zeta^2$. In practice, this shift distance ℓ_s will vary depending on the projection angle due to the change in scatter direction relative to the fixed coordinate system of the specimen. This can be compensated for, by rotating the coordinate system as

$$\ell_s = x_0 \cos\theta - y_0 \sin\theta, \tag{4.15}$$

where (x_0, y_0) are the coordinates of the centre of mass of the specimen and θ is the angle of the gantry.

To test if this model holds, we evaluated the change in scatter width and forward scatter factors for the narrow and broad components. The experiment consisted of the Monte Carlo simulation of a 10 cm slab of water shifted axially from its original position 50 cm away from the detector. We have plotted the resulting factors in Figure 4.4 for a 60 keV source, and have found the same trend holds throughout other energies.



(a) Shift amplitude factors

(b) Shift width factors

Figure 4.4: Shift amplitude and width factors for 10 cm water slab with 60 keV source.

It can be observed from Figure 4.4 that the amplitudes of both a narrow and broad Gaussian scatter components indeed fit $1/\zeta^2$ very well. The narrow Gaussian width factor also follows the expected value very well, though the broad factor appears to have a weaker dependence on shifts in the object. We note, however, that it empirically fits well to a modified factor of $\sqrt{\zeta}$, which we adopt in practice.

4.2.3 Compensation for Object Edges

A critical effect from inhomogeneous objects is the reduction of the scatter magnitude towards edges, as observed in [8, 128, 155]. This occurs since photons scattered away from an edge into a lower attenuating material such as air are far less likely to undergo further scattering events. Using a slab model such as SKS at the edges leads to an overestimation of scatter behind the thicker parts of the object, which will typically have the highest scatter to primary ratio. The consequences of this are not only that the attenuation through the line–of–sight measurements will be overestimated, but photon starvation and streak artefacts are likely to occur in the reconstruction [60] after its compensation.

In the fASKS work of [8], the authors observe an exponential drop in kernel magnitude towards an edge, which they approximate with the linear weighting factor $\boldsymbol{c}\{k\}(u,v)$ that can be applied between interfaces of thickness groups.

We employ a similar mechanism for use in our polyenergetic scatter model that instead produces a continuous modification of the kernel amplitude. From simulations on water slabs and elliptical tubes, we have found that the reduction in kernel amplitude is only significant on the broad Gaussian, which takes the following exponential form

$$\tilde{\boldsymbol{p}}_{\mathcal{B}} = \boldsymbol{p}_{\mathcal{B}} \exp\left(-\frac{\boldsymbol{t}_u^2 + \boldsymbol{t}_v^2}{c_{\mathcal{B}}^2}\right), \tag{4.16}$$

where $\boldsymbol{t}_u$ and $\boldsymbol{t}_v$ are continuous edge weights in the detector plane coordinates, and $\tilde{\boldsymbol{p}}_{\mathcal{B}}$ is the compensated broad forward scatter factor.

In Figure 4.5, we have shown the least squares fitting for $\boldsymbol{t}_u$ according to the model in (4.16). The data was generated from Monte Carlo simulations of an elliptical tube with semi-major axis of 15 cm semi-minor axis of 7.5 cm and length of 80 cm, where in each run the object was laterally shifted relative to the detector plane axis u. We are interested in the elliptical tube object since it crudely approximates the shape of a human body lying on a couch. With this, we also plotted the factor $\boldsymbol{c}\{k\}(u,v)$ as given in [8], though within the exponential model of (4.16).

From the nature of the kernel factors as observed in Figure 4.5, the increasing kernel attenuation factor towards the edge of the elliptical object can be seen. While this trend is somewhat caught by the linear edge factor, $\boldsymbol{c}\{k\}(u,v)$ from [8], it does not capture the

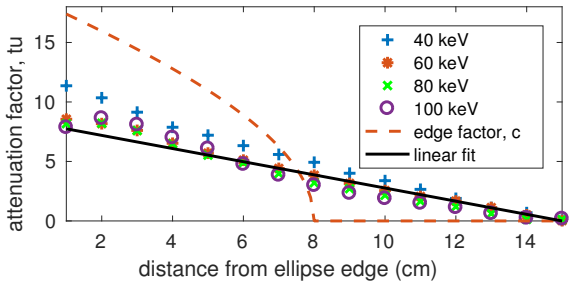


Figure 4.5: Broad kernel scatter shifts from simulated elliptical tube object fitted through nonlinear least squares at various source energies.

right functional form. Other observations from Figure 4.5 are that there is not a strong dependence on the incident energy, that there is a very small deviation from the slab model at the centre of the object, and the trend in $\boldsymbol{t}_u$ is approximately linear.

Motivated by this linear trend, we opt to use the following edge weighting function

$$\boldsymbol{t}_u\{k\}(u,v) = k_{\text{edge}}\boldsymbol{\tau}_e\{k\}(u,v)\frac{\partial \boldsymbol{\tau}_e\{k\}(u,v)}{\partial u}, \tag{4.17}$$

where $\boldsymbol{\tau}_e$ is the electron density projection as $\boldsymbol{\tau}_e = \boldsymbol{\Phi}\boldsymbol{\rho}_e$, and $\frac{\partial \boldsymbol{\tau}_e\{k\}(u,v)}{\partial u}$ is the spatial derivative of the electron density projection that can be calculated in practice as a discrete gradient. The factor k_{edge} is a scalar parameter to set the strength of the edge compensation effect. For isotropy, the factor $\boldsymbol{t}_v\{k\}(u,v)$ is calculated in the same way but with the derivative with respect to v.

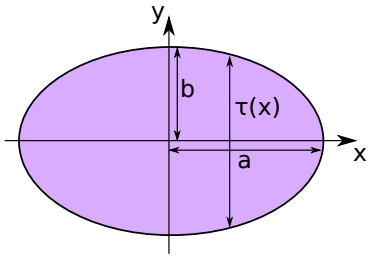


Figure 4.6: Diagram of ellipse for edge factor derivation.

It is easy to show that for a simple ellipse model, the scatter estimate in (4.17) gives the correct linear trend observed in Figure 4.5. Figure 4.6 shows an ellipse with major and minor semi axes lengths of a and b on x and y respectively. The thickness at a given point along x is denoted by $\tau(x)$. It can be shown that the variation in thickness with respect to x is expressed as

$$\frac{d\tau(x)}{dx} = -\frac{b^2 x}{a^2\tau(x)}. \tag{4.18}$$

Using the form of (4.17) in this case becomes

$$k_{\text{edge}}\tau(x)\frac{d\tau(x)}{dx} \propto \frac{b^2}{a^2}x, \tag{4.19}$$

which is now linear with the distance towards the edge, and dependent on the squared ratio between height and width of the ellipse. This second feature is also consistent with the experimental observation in [8], where they find an increase in edge effect with thicker objects.

To provide robustness of this method to rapidly varying projections, we found smoothing the electron density thickness $\boldsymbol{\tau}_e\{k\}(u,v)$ is beneficial. In the experimental sections, we employ a 2D Gaussian filter with standard deviation 1.5 cm. This smoothing also allows the edge compensation to be active towards sharp edges.

4.2.4 Complete Polyenergetic Scatter Estimation Model and Fitting

Combining the various elements, we note our complete scatter model — PolySKS — may be formed using the factors

$$\boldsymbol{S}_{\mathcal{N}}\{k\}(\omega_u,\omega_v) = \sum_{j=1}^{N_\xi} \mathcal{F}\left[\frac{1}{\zeta^2}\boldsymbol{p}_{\mathcal{N}}\{k\}(u,v,\xi_j)\right] \odot \mathcal{F}\left[\exp\left(-\frac{\boldsymbol{u}^2+\boldsymbol{v}^2}{\zeta^2 c^2_{\mathcal{N}}(\xi_j)}\right)\right], \tag{4.20}$$

and

$$\begin{aligned}\boldsymbol{S}_{\mathcal{B}}\{k\}(\omega_u,\omega_v) =& \mathcal{F}\left[\frac{1}{\zeta^2}\sum_{j=1}^{N_\xi}\boldsymbol{p}_{\mathcal{B}}\{k\}(u,v,\xi_j)\exp\left(-\frac{\boldsymbol{t}_u^2+\boldsymbol{t}_v^2}{c^2_{\mathcal{B}}}\right)\right] \\ &\odot\mathcal{F}\left[\exp\left(-\frac{\boldsymbol{u}^2+\boldsymbol{v}^2}{\zeta c^2_{\mathcal{B}}}\right)\right].\end{aligned} \tag{4.21}$$

and as before the estimate is then

$$\boldsymbol{s}\{k\}(u,v) = \mathcal{F}^{-1}\left[\boldsymbol{S}_{\mathcal{N}}\{k\}(\omega_u,\omega_v) + \boldsymbol{S}_{\mathcal{B}}\{k\}(\omega_u,\omega_v)\right] \text{ for } i = 1,...,N_{\text{proj}},. \tag{4.22}$$

In calculating (4.20), the exact form of forward scatter factor used is

$$\boldsymbol{p}_{\mathcal{N}}\{k\}(u,v,\xi_j) = K_{\mathcal{N}}(\xi_j)\boldsymbol{b}\{k\}(u,v,\xi_j) \odot \exp(-[\boldsymbol{\Phi}\boldsymbol{\mu}(\xi_j)]\{k\}) \odot [\boldsymbol{\Phi}\boldsymbol{\rho}_e]\{k\}, \tag{4.23}$$

whilst the broad component of (4.21) utilises the fitting parameters as

$$\boldsymbol{p}_{\mathcal{B}}\{k\}(u,v,\xi_j) = K_{\mathcal{B}}(\xi_j)\boldsymbol{b}\{k\}(u,v,\xi_j) \odot \exp(-h_1(\xi_j)[\boldsymbol{\Phi}\boldsymbol{\mu}(\xi_j)]\{k\} + h_2(\xi_j)\log([\boldsymbol{\Phi}\boldsymbol{\rho}_e]\{k\})). \tag{4.24}$$

In order to use this model, one requires to find the energy dependent parameters c_{N}, $K_{\mathcal{N}}$, $K_{\mathcal{B}}$, h_1 and h_2, as well as the energy independent parameters c_{B} and k_{shift} from (4.17) to calculate $\boldsymbol{t}_u$ and $\boldsymbol{t}_v$. Apart from the edge factor parameters, these were fitted to data from simulating scatter through water slabs having a thickness range of 1 cm to 40 cm with 1 cm increment, and for a number of different photon energies. In each case, we used a detector response function derived from the energy absorption coefficient of Cesium Iodine (CsI) with a thickness of 0.6 mm to replicate a photo absorption scintillator, and a detector of 40×40 cm with resolution 256×256. We then minimised the nonlinear least squares error against this data to our model using Newton's method, assuming a fixed broad width of 35 cm. These parameters are shown in Table 4.1.

Table 4.1: PolySKS parameters from water kernel fitting used throughout the experimental section.

energy (keV)	$K_{\mathcal{N}}$	$c_{\mathcal{N}}$(cm)	$K_{\mathcal{B}}$	h_1	h_2
40	1.38×10^{-6}	6.91	3.51×10^{-7}	0.876	1.10
60	1.40×10^{-6}	4.77	3.16×10^{-7}	0.828	1.15
80	1.39×10^{-6}	3.62	2.97×10^{-7}	0.804	1.20
100	1.40×10^{-6}	2.90	2.79×10^{-7}	0.791	1.25
120	1.40×10^{-6}	2.44	2.66×10^{-7}	0.785	1.29

Table 4.2: Energy independent parameters used for PolySKS fitting.

$c_{\mathcal{B}}$(cm)	k_{edge}(full fan)	k_{edge}(half fan)
35	2.35	1.57

From Table 4.1, one can see typical parameters from our fitting. It is interesting to see that although we included the energy dependence on $K_{\mathcal{N}}$ to account for finite detector effects, there is almost no dependence from our fitting, which agrees with theory from [40]. Also interesting is the significant deviation of the narrow width and broad forward model parameters throughout energy. A critical factor to consider is that the scatter scaling factors $K_{\mathcal{N}}$ and $K_{\mathcal{B}}$ will be dependent on the area of the detector, and they should be appropriately scaled against our detector area of 0.0244 cm^2 for other resolutions.

For the perturbation factor k_{edge}, we calculated the average factor across energies for the elliptical tube experiment presented in Figure 4.5 according to the model in (4.17). When we repeated this for an offset detector, we found a reduction of around 33%. This may be due to wider cone angles of the geometry having less of an effect from edges perpendicular to the detector. For the other parameters however, we found they hold in both full fan or half fan acquisitions.

The energies used in Table 4.1 correspond to those we simulated in Monte Carlo. In the following experimental sections of this chapter, we use 21 energies in the range 20–120 keV. To find appropriate parameters for all these energy sample points, we interpolate the values given in Table 4.1, using a shape-preserving piecewise cubic interpolation. For the lowest energies in the range 20–40 keV, we find they contribute a negligible amount of scatter, but we use extrapolation of the parameters to obtain values for these nonetheless.

For the Polyquant parameters, we used those given in Appendix A, for both numerical and physical phantom experiments.

4.3 Incorporating the Polyenergetic Scatter Model in Polyquant

Although the polyenergetic model as derived in Section 4.2 is a general model for scatter that may be used in various algorithms or applications, its exact formulation works particularly well in the Polyquant framework, Polyquant–PolySKS. To derive an algorithm, we start with the gradient term of the negative log-likelihood NLL function with respect to the electron density, which has the parameter

$$\boldsymbol{d}(\boldsymbol{\rho}_e) = \boldsymbol{y} \oslash \left(\sum_{j=1}^{N_\xi} \boldsymbol{\psi}(\boldsymbol{\rho}_e, \xi_j) + \boldsymbol{s}(\boldsymbol{\rho}_e) \right) - \mathbf{1}, \tag{4.25}$$

where $\boldsymbol{s}(\boldsymbol{\rho}_e)$ is the updated scatter estimate based on the current iterate using (4.23), and $\boldsymbol{\psi}(\boldsymbol{\rho}_e, \xi_j)$ is defined in (3.8). This is then used to calculate the derivative as

$$\boldsymbol{\nabla}\text{NLL}(\boldsymbol{\rho}_e; \boldsymbol{y}) \approx \sum_{i=1}^{N_f} \boldsymbol{f}_i(\boldsymbol{\rho}_e) \odot \boldsymbol{\Phi}^T \left[\sum_{j=1}^{N_\xi} \alpha_i(\xi_j) \boldsymbol{\psi}(\boldsymbol{\rho}_e, \xi_j) \odot \boldsymbol{d}(\boldsymbol{\rho}_e) \right], \tag{4.26}$$

in the same manner as the regular Polyquant method [149]. We note a subtle difference is the dependence of the scatter term with $\boldsymbol{\rho}_e$. With this, the exact formulation of the gradient would include a term from the derivative of the scatter w.r.t. the current estimate, which for convolutional models would be an extra filtering term. We have not included this in (4.26), since we have found it only has a weak dependence on the NLL and hence can be ignored. For these reasons, we have ignored this dependence and have symbolised this non-exactness by the '$\approx$' sign.

Calculating the term $\boldsymbol{s}(\boldsymbol{\rho}_e)$ requires from (4.5) the terms $\boldsymbol{\Phi}\boldsymbol{\mu}(\xi)$ for each energy and $\boldsymbol{\Phi}\boldsymbol{\rho}_e$. For the first term, we note from the Polyquant formulation in (3.2) we have

$$\boldsymbol{\Phi}\hat{\boldsymbol{\mu}}(\xi_j) = \sum_{i=1}^{N_f} \alpha_i(\xi_j)\boldsymbol{\Phi}[\boldsymbol{f}_i(\boldsymbol{\rho}_e) \odot \boldsymbol{\rho}_e] + \beta_i(\xi_j)\boldsymbol{\Phi}\boldsymbol{f}_i(\boldsymbol{\rho}_e), \tag{4.27}$$

where the expensive factors to compute, $\boldsymbol{\Phi}[\boldsymbol{f}_i(\boldsymbol{\rho}_e) \odot \boldsymbol{\rho}_e]$ and $\boldsymbol{f}_i(\boldsymbol{\rho}_e)$, are already available at each iteration. Additionally, due to the properties of the threshold functions given in (3.3), we can find that

$$\boldsymbol{\Phi}\boldsymbol{\rho}_e = \sum_{i=1}^{N_f} \boldsymbol{\Phi}[\boldsymbol{f}_i(\boldsymbol{\rho}_e) \odot \boldsymbol{\rho}_e], \tag{4.28}$$

which also only requires projections that are already available at each iteration anyway.

4.4 Evaluation from Simulation

In the first experiment, we evaluate the accuracy of our proposed integrated scatter model, Polyquant–PolySKS, with comparison to precomputed scatter. To synthesise the data, we performed a Monte Carlo simulation of a CBCT system using the software Gate [26] according to the method outlined in [140].

We generated two acquisitions of the female ARCP [41] from the pelvis and head region, using half and full fan scans respectively. In both cases, we used a planar detector of 40×30 cm, placed 150 cm from the source and 50 cm from the centre of rotation. For the half fan pelvis case, we offset the detector by 16 cm and used a shaped source to replicate the effect of a bow-tie filter. We also used a appropriate source profile for the full fan head case. For both scans, we simulated a total of 1×10^{11} photons over 160 projections and with a detector resolution of 256×128. The image resolutions were those of the original

ARCP at 299×137 and 60 slices of 0.484 cm thickness. We do not include any scatter grid in our simulation.

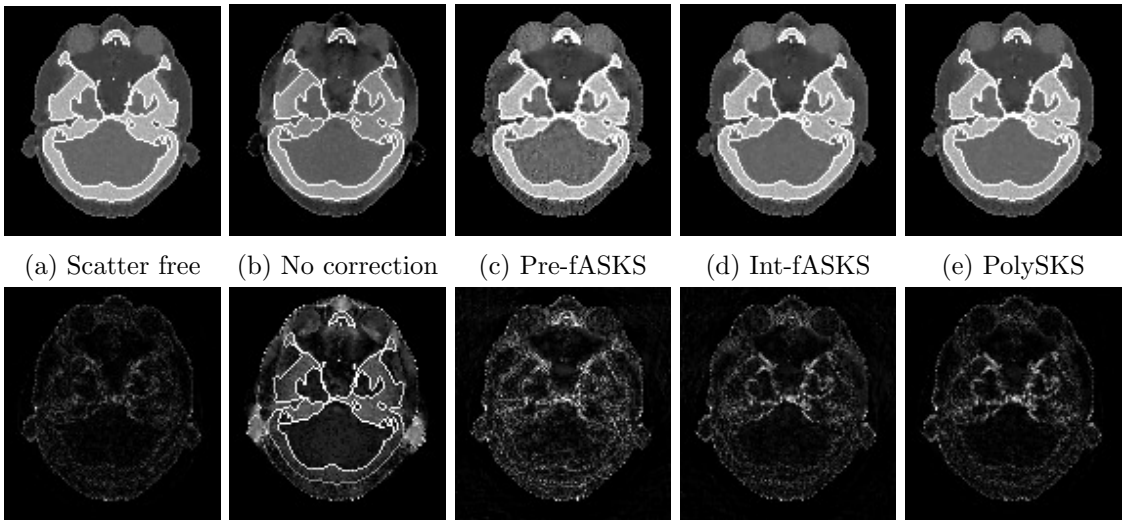
All methods under test use the Polyquant reconstruction technique [149], with identical TV regularisation, allowing a fair comparison between scatter estimation approaches. We will thereby refer to the reconstruction method Polyquant–PolySKS only by the scatter estimation method PolySKS. The spectra used were equivalent to 100 kVp and 120 kVp tube potentials for the head and pelvis cases respectively, which we equally sampled into 21 energies. The spectra were generated using the SpekCalc software [143] for a tungsten anode at 30°. The system operators $\mathbf{\Phi}$ and $\mathbf{\Phi}^T$ were implemented as separable footprint projections [70] within the Michigan Image Reconstruction Toolbox in MATLAB. The methods we will use for comparison are:

- *Scatter free*: we have entirely removed all scatter from the numerical data, which represents the gold standard of any scatter mitigation method.
- *No correction*: scatter is included in the measurements, but no attempt is made to correct it, to show the extent of scatter induced artefacts and represents the 'worst case'.
- *Pre-fASKS*: Polyquant with pre-calculated fASKS calculated according to (2.73), applied iteratively to the measurements 10 times. For fairness, we refit all parameters to the same simulated water slabs as were used to inform the PolySKS parameters in Table 4.1. For the perturbation factors, we used the same edge factor as presented in the original publication [8], and used $\gamma = 0.04$ and $\gamma = 0.03$ for the head and pelvis cases respectively, selected to give the best fitting against our elliptical shift data used to fit k_{edge}.
- *Int-fASKS*: an identical scatter model and parameters as in *Pre-fASKS*, but performed at each iteration using the forward scatter factor (2.70) in place of its approximation in (2.71). This does not exploit either the spectral information nor the electron density.
- *PolySKS*: our proposed polyenergetic integrated scatter model as detailed in Section 4.2.1 to give Polyquant–PolySKS, according to parameters given in Table 4.1.

For each method, we performed 10 epochs in the head case and 20 epochs for the pelvis. It seemed the lower amount of attenuating material in the head case meant these methods

convergence significantly faster.

To evaluate the performance of the methods, we have shown reconstructed images in Figures 4.7 and 4.8. We also quantified the accuracy through the root–mean–squared–error (RMSE) through slices 18–42 — corresponding to slices in the cone angle — and these are tabulated in Table 4.3.



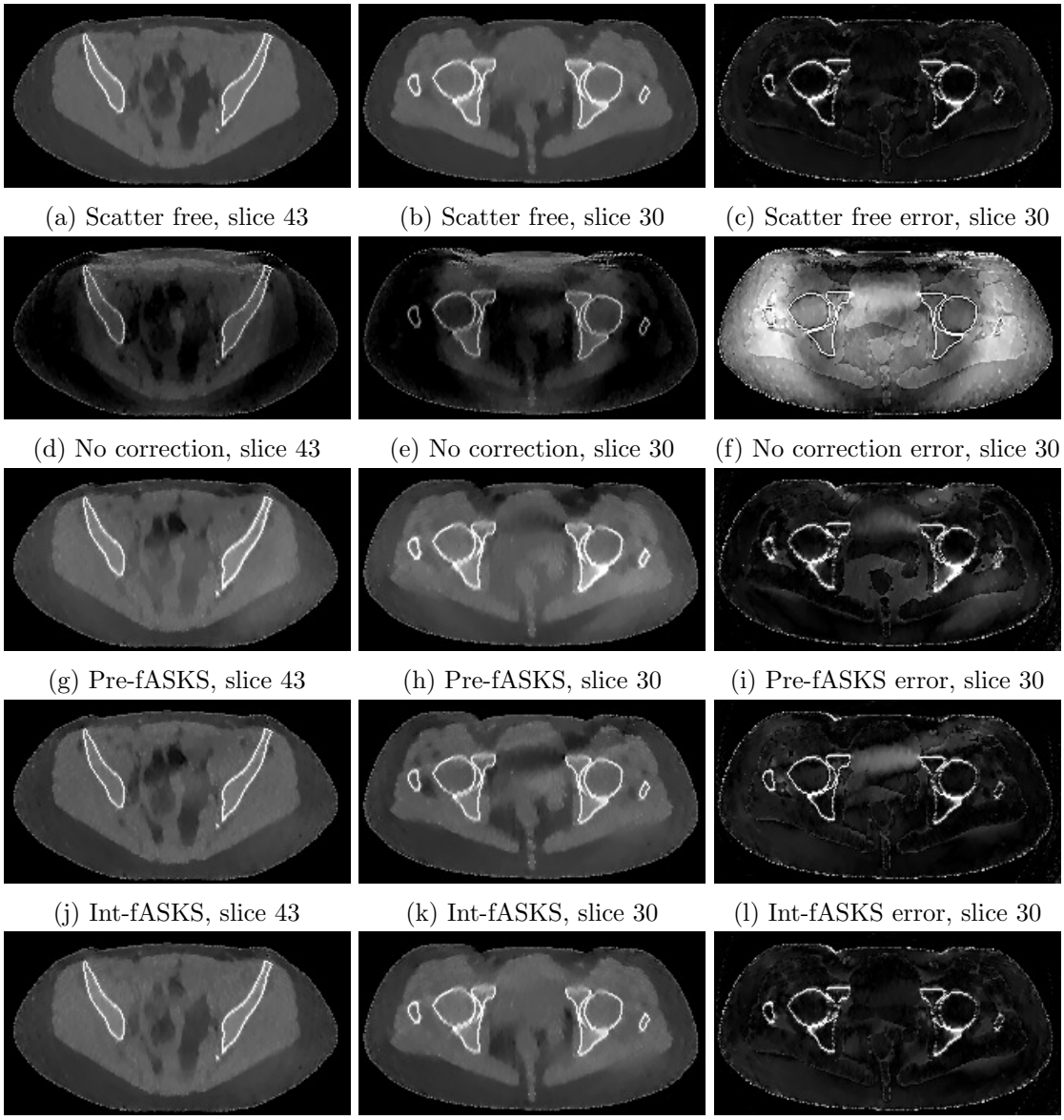
(a) Scatter free (b) No correction (c) Pre-fASKS (d) Int-fASKS (e) PolySKS

(f) Error from (a) (g) Error from (b) (h) Error from (c) (i) Error from (d) (j) Error from (e)

Figure 4.7: Visual results of slice 36 from the head CBCT data with display window [0.8,1.4] to aid visualisation of soft tissue and reconstruction artefacts; absolute errors have display window [0,0.3].

For the scatter contaminated head phantom reconstruction, shown illustratively in Figure 4.7b and 4.7g, the scatter leads to visible shading artefacts, especially around the edges and close to the bones, as well as a significant underestimation in the bony structures. Of the scatter correction methods, both the integrated scatter estimation approaches Int-fASKS and PolySKS perform similarly, whilst the Pre-fASKS still exhibits shading artefacts and has a higher level of noise. Quantitatively, from the second column in Table 4.3, this trend is replicated, with all correction methods considerably closing the gap between 'scatter free' and 'no estimate', with the PolySKS giving the highest accuracy. The difference between the fASKS methods is especially interesting as it implies one must incur some loss from the deconvolution and using (2.71) in place of (2.70).

The pelvis reconstructions are then shown in Figure 4.8. Here, scatter is much more of a critical factor, with the 'no estimate' in Figures 4.8d–f showing huge errors throughout the image. In this case, both Pre-fASKS and Int-fASKS exhibit strong shading artefacts also, especially in the region between the bones. Although PolySKS does have a number



(a) Scatter free, slice 43 (b) Scatter free, slice 30 (c) Scatter free error, slice 30

(d) No correction, slice 43 (e) No correction, slice 30 (f) No correction error, slice 30

(g) Pre-fASKS, slice 43 (h) Pre-fASKS, slice 30 (i) Pre-fASKS error, slice 30

(j) Int-fASKS, slice 43 (k) Int-fASKS, slice 30 (l) Int-fASKS error, slice 30

(m) PolySKS, slice 43 (n) PolySKS, slice 30 (o) PolySKS error, slice 30

Figure 4.8: Visual results of slices 43 and 30 from the pelvis CBCT data with display window [0.8,1.4] to aid visualisation of soft tissue and reconstruction artefacts; absolute errors have display window [0,0.3].

of overestimated regions in the fat regions visible in the error plot in Figures 4.8o, it is generally a good match to the 'scatter free' reconstruction. Quantitatively from Table 4.3, the same trend as the head case can be seen, with the Int-fASKS giving roughly a 9% drop in error over Pre-fASKS, and PolySKS giving another 9% reduction over that.

Table 4.3: Quantitative results: relative electron density (ρ_e) RMSE of slices 18–42.

Scheme	Head RMSE	Pelvis RMSE
Scatter free	0.0112	0.0322
No estimate	0.0413	0.122
Pre-fASKS	0.0258	0.0479
Int-fASKS	0.0183	0.0435
PolySKS	**0.0152**	**0.0398**

4.5 Validation from Real Data

In this section, we tested our method on real data from CBCT scans of physical phantom objects. For this, we have two quantitative CIRS (Computerized Imaging Reference Systems, Inc., Norfolk, VA (USA)) phantoms with acquisitions on a Varian TrueBeam On-Board Imager (Varian Medical Systems, Inc., Palo Alto, CA (USA)).

In each of the two phantom studies we compare these three approaches:

- *No estimate*: Polyquant reconstruction with no scatter estimate. This is included to indicate the magnitude of the scatter artefacts and to represent a worst case.
- *Pre-fASKS*: Polyquant reconstruction is performed using the estimate of scatter generated by the TrueBeam system as preprocessing. This is an instance of fASKS shown in (2.73), but with the manufacturers parameters, and with a constant tailed triangle modulation of the kernel shapes to account for the scatter grid on the scanner, the inclusion of a detector glare model [8] and a heuristic to improve upon scatter inaccuracies from the couch as in [156].
- *PolySKS*: our proposed integrated quantitative scatter model as outlined in Section 4.2.1. In this case, we use the very same model and parameters from Table 4.1 as used in the simulated section, though with the inclusion of a similar modulation for the presence of a scatter grid as with Pre-fASKS.

For every method, we ran Polyquant for 50 epochs. Unlike the simulated experiments, we did not evaluate the Int-fASKS approach since we did not have means to modify the system's scatter approach, and we have already demonstrated the superiority of PolySKS over this in the numerical results.

4.5.1 CIRS Insert Phantom

In the first case, we investigate reconstructions of a CIRS electron density insert phantom. The acquisition consisted of 896 projections with a 100 kVp tube potential and a current of 20 mA over a 15 ms pulse duration.

The phantom itself has a cross sectional diameter of 18 cm and a thickness of 5 cm. This poses a challenge for scatter estimation due to its thinness, since it naturally deviates significantly from the underlying assumed slab model.

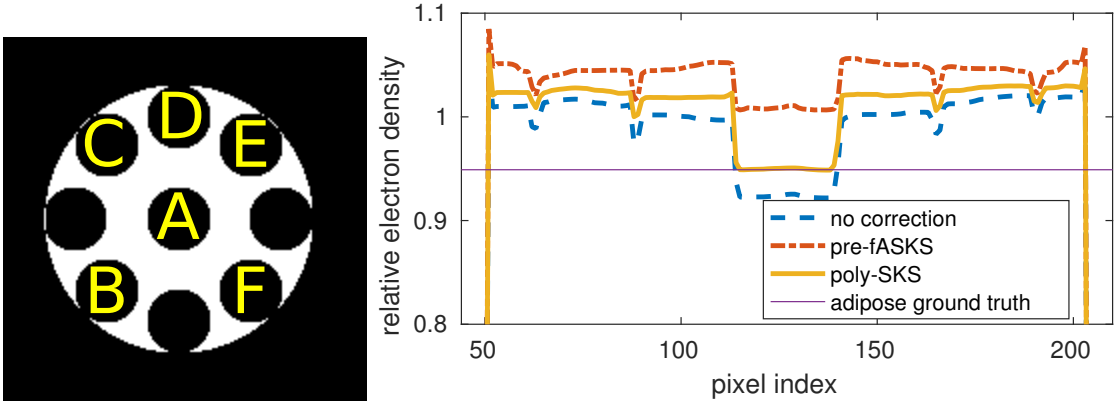
Table 4.4 shows material information for the tissue inserts in the phantom. Of particular interest is the relative electron densities ρ_e, since these will act as the ground truths against which to compare our quantitative estimates. To calculate the estimates from the reconstructions, we took the mean value from a circular region of radius 1 cm of each insert. Both the body of the phantom and the central left and right inserts are manufactured from 'water equivalent' resin. Since we are pursuing the quantitatively accurate estimation of soft tissues, we have not evaluated these. However, since we would like to investigate the shading effect of scatter on homogeneous slabs, we do assess the deviation with the body structure.

Table 4.4: Quantitative results: relative electron density accuracy of insert phantom.

Label	A	B	C	D	E	F
Material	Adipose	Muscle	Soft bone	Muscle	Breast	Liver
Density (g/cm^3)	0.96	1.06	1.16	1.06	0.99	1.07
RED, ρ_e	0.949	1.043	1.117	1.043	0.949	1.052

(a) No estimate (b) Pre-fASKS (c) PolySKS

Figure 4.9: Illustrations of reconstructed volumes of insert phantom of central slice with display window [0.8,1.2].



(a) Regions of interest

(b) Profiles through centre of reconstructions

Figure 4.10: Insert location labels, along with slab ROI mask, and line profiles through centre of reconstructions.

Reconstructions from the insert phantom are shown in Figure 4.9. To supplement this, we have also plotted a line profile horizontally through their centres in Figure 4.10. From observation, all three methods produce similar reconstructions. In the case of no estimate in Figure 4.9a, there is a visible shading towards the centre of the object also visible in its profile, which is very typical of a scatter artefact. Whilst both the Pre-fASKS and PolySKS compensate for this, PolySKS appears to generate more of a uniform slab estimation, with the former producing a raised estimate at the centre and edges of the profile in Figure 4.10. This overestimation is likely caused by overestimating scatter, perhaps arising the thinness of the phantom.

Table 4.5: Quantitative results: relative electron density accuracy of STEEV phantom.

	Tissue accuracy (% error)						Slab deviation (std)
Scheme	A	B	C	D	E	F	ROI
No estimate	−2.73	−1.08	**1.72**	−1.04	**1.70**	−1.75	7.74×10^{-3}
pre-fASKS	6.23	2.32	8.01	2.14	5.07	1.69	5.38×10^{-3}
PolySKS	**0.0718**	**0.390**	3.91	**0.171**	3.56	**−0.277**	$\mathbf{4.77 \times 10^{-3}}$

Table 4.5 then shows the quantitative analysis of the reconstructions. It is clear that Pre-fASKS produces an overestimation of the electron densities across the board, which in many cases represents a worse performance than from no estimate. PolySKS on the other hand performs very well in most cases, having an error within the stated material manufacturing tolerance of 1%, with the exception of the breast and bone materials. Although it may appear disconcerting that the no estimate has the most accurate

performance in the bone and breast cases, we note that these are still overestimations. Since in reality there will be some non-zero scatter behind these structures, and the addition of any scatter compensation will raise the estimation of density, it is unsurprising that both Pre-fASKS and PolySKS overestimate this further, so we suggest this is unlikely arising from a deficit in scatter estimation. Possible causes for the discrepancy may be inaccurate spectrum modelling or effects from objects outside the field of view such as the couch.

The other reported value in Table 4.5 is the slab deviation, which is the standard deviation of all pixels within the ROI in Figure 4.10a. Here, we note that either scatter modelling strategy offers a large increase in uniformity by at least 30% over 'no estimate', with a further 11% increase from Pre-fASKS to PolySKS. This result is unsurprising considering the flatness of the profiles in Figure 4.10.

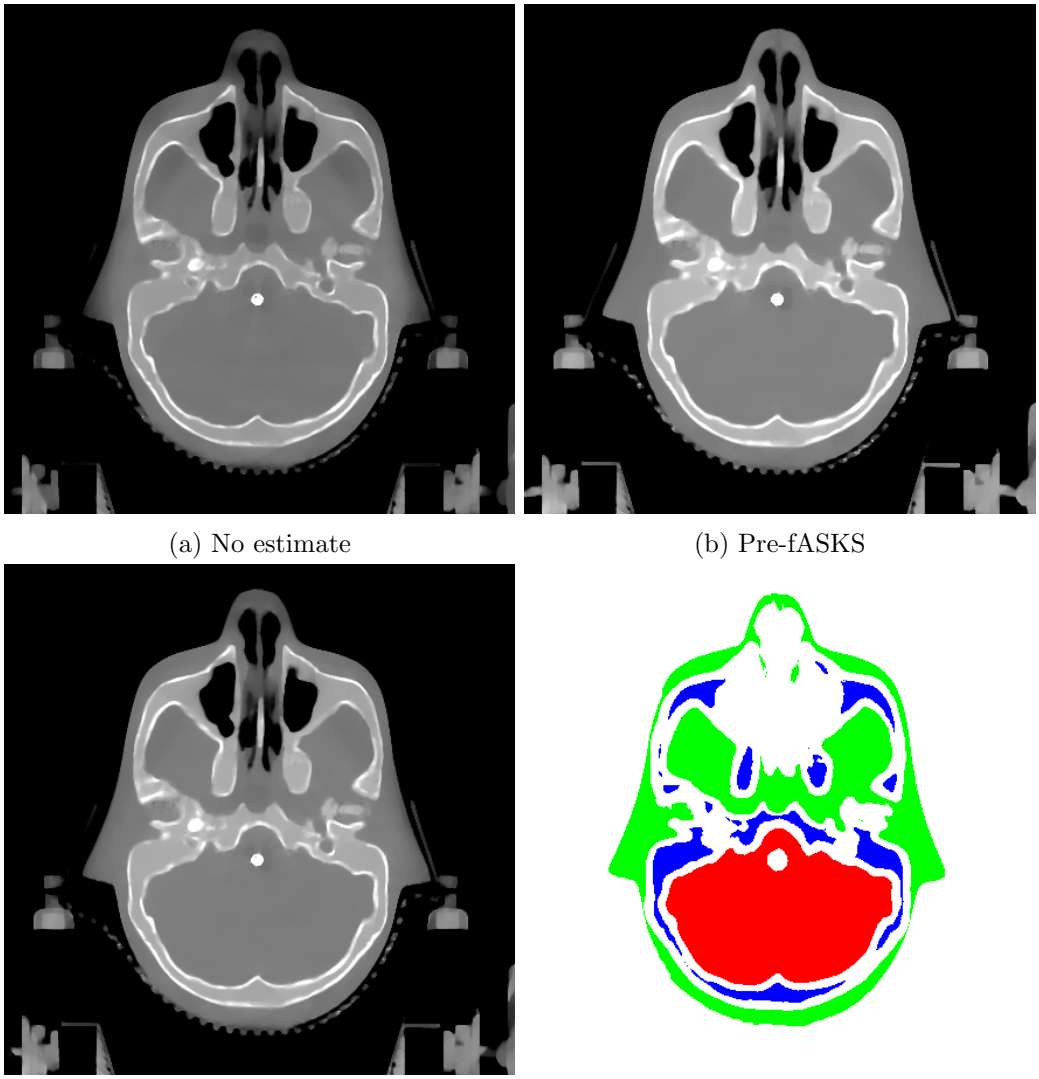
In summary from this experiment: PolySKS results in the most uniformity throughout homogeneous structures, and is the only variant offering highly quantitatively electron density estimation, apart from the bone and breast that are overestimated in every case.

4.5.2 STEEV Head Phantom

In the second case, we performed reconstruction of the CIRS STEEV head phantom. In this case, the phantom consists of complex resin structures to mimic the attenuation and form of a human head. Again, we wish to analyse the quantitative accuracy and uniformity of the homogeneous materials.

Reconstructions of the head phantom are shown in Figure 4.11. Observations are that from the 'no estimate' in Figure 4.11a exhibits similar shading artefacts in the soft tissue regions as the simulated head case in Figure 4.7b. Between the two scatter estimation approaches, both appear to similarly correct for inhomogeneities from scatter, though the Pre-fASKS does show higher intensities than PolySKS. Although we are unsure of the exact reason why fASKS leads to a consistent overestimation of scatter in our real CBCT data experiments, we should stress that the method is designed only to enhance uniformity, subject to a subsequent calibration of the image, unlike our approach of direct quantification.

To quantitatively assess these reconstructions, we calculated the mean and standard



(a) No estimate (b) Pre-fASKS

(c) PolySKS (d) Tissue ROIs

Figure 4.11: Illustrations of reconstructed volumes of STEEV phantom with display window using slice 77 [0.7,1.4].

deviation of homogeneous tissue types: brain, soft tissue and soft bone. We isolate regions from these materials with a simple threshold based segmentation of the Pre-fASKS reconstruction, and have shown the coloured regions in green, red and blue respectively in Figure 4.11d. These are then reported in Table 4.6.

From the quantitative results in Table 4.6 it can be seen that similarly to the insert phantom, the PolySKS scatter integration leads to the most quantitatively accurate reconstruction. Also similar to the insert case, is the trend with the no estimate leading to an underestimation of electron density, and Pre-fASKS overcompensating to an over estimation. With the exception of the RMSE on the soft tissue, the PolySKS is the best performing method in every metric. In the soft tissue RMSE case, the Pre-fASKS is the

Table 4.6: Quantitative results on soft tissues slices 77.

Material	Soft Tissue		Brain		Soft Bone		Total
ROI colour	green		red		blue		all
ρ_e	1.028		1.039		1.157		–
	% mean	RMSE	% mean	RMSE	% mean	RMSE	RMSE
No estimate	−2.26	0.0430	−1.41	0.0158	−3.14	0.0409	0.0345
Pre-fASKS	1.44	**0.0201**	1.57	0.0173	3.92	0.0478	0.0253
PolySKS	**0.0544**	0.0209	**0.439**	**0.00709**	**0.687**	**0.0169**	**0.0162**

best performing approach due to a lower variation throughout this tissue. It could be that one of the additional correction factors such as the asymmetric modulation, detector glare or couch scatter, not included in our tested PolySKS implementation, may be worthwhile including in our approach also. We also expect the parameters in Table 4.1 may be improved through fitting to real measurements of scatter. Even with this discrepancy though, PolySKS shows a significant advantage over the whole image, with a 36% drop in total RMSE over Pre-fASKS.

4.6 Discussion

There a several important aspects of our method, about which we would like to discuss: its extension with other correction factors and models, the computational impact of the new scatter model, and its use in other models besides Polyquant.

4.6.1 Extension of the PolySKS Model

Due to the non-exactness of convolutional scatter estimations, there are many heuristics that may enhance their practical performance under certain conditions. In our presentation of PolySKS, we have tried to present a simple model, whilst employing a couple of empirical perturbations that were found to give significant performance gains for minimal computational cost. Naturally, one could employ several other extensions also, and we will discuss some of these options.

Firstly, we have not included the asymmetric modulation of fASKS in our model — active when $\gamma > 0$ in (2.73). This could certainly be applied both on the total scatter factors, or

separately on each energy signal. We have performed preliminary experiments in both of these cases, but did not find any significant gain in performance for our numerical data, whilst increasing the computational cost by at least $2\times$. Instead, we found just the edge factor derived in Section 4.2.3 gave good compensation of inhomogeneities with our data. It could be that objects with more complicated variations in thickness such as the torso, with arms and lung regions may benefit from its inclusion.

Another factor that may be worth considering is the effect of objects outside the field of view, such as the patient couch or scatter from other hardware. In [156], the authors do report on this having significant effects on scatter estimation, and its inclusion may widen the gap between PolySKS and the commercial fASKS, where this is employed, in our experimental section. In our numerical tests however, we do not have any additional hardware, allowing us to fairly characterise the relative effectiveness of the methods under test.

We have focussed on convolutional scatter estimation due to their quick run time with FFT implementations. However, a new method published during the preparation of this thesis, using the linear Boltzmann transport equation, promises both speed and accuracy [160, 161]. Due to the wealth of information available during the iterates of Polyquant, we believe this method may be readily integrated into the algorithm as with Polyquant–PolySKS, but the gain this may offer remains to be demonstrated. In any case, it has been shown that a similar model may be used to perturb convolutional models and increase their accuracy in [157], which may also offer benefit to our method.

Another potential avenue of research is attempting to apply convolutional neural network learning to our polyenergetic scatter factor derived in this chapter. A similar approach was published by [139] during the development of this thesis, but this was a preprocessing correction for monoenergetic assumptions. It is possible that training a network with our integrated framework would build in an explicit independence to source potential, making it much more universal.

4.6.2 Computational Feasibility

A critical issue with scatter estimation methods is their computational complexity, especially when they are performed at every iteration of reconstruction, such as with

Polyquant–PolySKS. Since the motivation behind maintaining a spatially invariant convolutional approach is fast implementation, it is certainly worth directly addressing this point.

As we highlighted in Section 4.2.1, calculating the convolution itself bears a cost of N_{proj} IFFT operations and $N_{\text{proj}}(N_\xi + 1)$ FFT operations per data pass. Comparatively, the fASKS method has a convolution cost of $2N_{\text{proj}}N_{\text{iter}}$ IFFTs and $2N_{\text{proj}}N_{\text{thick}}N_{\text{iter}}$ FFTs, where N_{thick} are the number of thickness groups and N_{iter} are the number of iterations, which are set to 3 and 4 respectively in [8]. If one set $N_{\text{thick}} = 3$ and $N_\xi = 21$ as we have in our experiments, then we note that one could only perform 1.4 epochs for the same cost of fASKS. Therefore, for the 50 data passes we use in our real data experiments, PolySKS carries a notably higher cost. However, we believe that sub-sampling the energy bins is unlikely to have a vast impact of its accuracy, whilst providing big speed ups.

Although when compared to fASKS our estimation approach is likely to be more expensive, we highlight the fact that relative to the cost of performing iterative reconstruction, it is still cheap. In the STEEV head phantom experiment for instance, the PolySKS result came at only a 20% higher cost than Pre-fASKS, with no attempt to sub-sample any of the scatter calculation. Our implementation made in MATLAB is also likely to be sped up significantly by rewriting in a compiled language such as C++, for which [8] saw a 4.7× speed up for a single thread with their fASKS — an implementation on GPGPU hardware is likely to be many times faster still. Additionally, over methods such as Monte Carlo or single scatter models, PolySKS is still likely to be many times faster.

4.6.3 Application of PolySKS in Other Methods

Although we have developed this scatter model specifically for the Polyquant framework, it is certainly available for alternative methods also. Firstly, due to the reported advantage of Int-fASKS over Pre-fASKS, one may use this scatter integration concept with any iterative method, and expect to realise a significant gain by also fusing the scatter estimation with reconstruction.

With the PolySKS scatter model, we believe it may be incorporated into most polyenergetic likelihood models, though with concessions. For methods that quantify a monoenergetic attenuation or mass density rather than electron density such as IMPACT

[11] or Poly-SIR respectively [46], the information $\mathbf{\Phi}\boldsymbol{\rho}_e$ is not available at each iteration. To overcome this, one could either perform a non linear conversion from the projection information $\mathbf{\Phi}\boldsymbol{\mu}(\xi)$, or indeed use directly $\mathbf{\Phi}\boldsymbol{\mu}(\xi)$ in place of $\mathbf{\Phi}\boldsymbol{\rho}_e$. This would in fact be directly proportional for water, with the constant of proportionality being its mass attenuation coefficient, but is likely to lose accuracy over diverse material classes.

As a preprocessing method before an analytic reconstruction such as FDK [9] however, the polyenergetic convolutional model can not be directly applied, due to not having an energy decomposition of the attenuation available. One possible way to exploit it though may be to use an approximate water–bone beam hardening model such as [45] to approximate the spectral properties. Additionally, from dual energy (or spectral) CT measurements, such a decomposition may be inferred and the model may be applicable. In both cases, one would likely have to apply the model iteratively as a deconvolution of the contaminated primary measurements as in Pre-fASKS [8] and it is not clear this would offer any gain.

4.7 Conclusions

We have introduced a polyenergetic convolutional scatter model, and integrated it into a quantitative iterative reconstruction method, denoted Polyquant–PolySKS. Additionally, have developed a fast algorithm for its implementation. The advantages of this are twofold: we have demonstrated through using the same convolutional model, fASKS, that one achieves better reconstruction performance when integrated rather than used in preprocessing; and we have demonstrated that one can exploit polyenergetic information available during Polyquant to gain a higher accuracy still. Due to its spatial invariance, the scatter calculation can be done rapidly through FFT operations, and contributes only a fraction of additional computational time over an iterative reconstruction. From our experiments with both numerical and real data, we have demonstrated superior electron density accuracy with this method, which promises to enhance quantitative medical procedures such as radiation therapy dose calculation.

Chapter 5

Quantitative Imaging for Dose Calculation

5.1 Introduction

Throughout this work, we have pursued quantitative imaging as a means to enhance radiotherapy planning accuracy. In this section, we will put these ideas to the test, and evaluate how our developed imaging techniques translate to dose prediction enhancements.

As outlined in the background chapter in Section 2.1, quantifying CT reconstructions is an integral part of the radiotherapy workflow. However, the accurate quantification from CBCT is challenging due to its sampling deficit, the desire to minimise dose, and from high levels of scatter. Several attempts to perform accurate CBCT either from reconstruction [162] [163], registration [164] or a combination of the two with PIRPLE [107] [108] have been carried out, although these are substantially different from the direct quantitative methods presented in this thesis, and dose accuracy was not explicitly evaluated. In [21] the authors do actively evaluate the dose planning accuracy from scatter corrected CBCT, though the reconstruction is based on the FDK method in that case.

In this study, we will present a comprehensive evaluation of the dose planning potential from CBCT reconstruction. Moreover, we show how our develop methods can translate their enhanced electron density accuracy into a tangible enhancement of dose calculation.

From a medical perspective, the important factors for analysing dose prediction, are how it may translate into enhanced efficacy and reduced incidences of complications. In 2010, the quantitative analysis of normal tissue effects in the clinic (QUANTEC) was published [165] [166]. This work reviewed hundreds of clinical trials, to produce a quantitative summary of toxicity risks to many healthy organs, from an associated dose from radiotherapy. Throughout the experimental sections, we will be analysing the

dose error in various scenarios, and will comment on potential implications of these with reference to QUANTEC in the discussion section.

In the first set of experiments, we will perform numerical simulations for a number of different anatomical sites and radiation therapy types: photons, protons and carbon ions. In this numerical setting, we evaluated several competing reconstruction methods, as well as the full Polyquant–PolySKS as presented in Chapter 3. Secondly, we will test the dose calculation accuracy against physical energy absorption measurements on the STEEV head phantom irradiated with a commercial and clinical photon therapy linear accelerator. We performed this real dose validation work prior to finalising the scatter method in Chapter 3. For this reason, we have not evaluate Polyquant–PolySKS here, though this is an area of future research.

5.2 Numerical Evaluation of Radiation Planning

The numerical dose planning evaluation study was performed in the radiation therapy planning suite matRad [35] from the German Cancer Research Centre (Deutsches Krebsforschungszentrum — DKFZ), which is a freely available package for MATLAB allowing the calculation of treatment plans through different types of radiation.

For the data, we utilised the synthetic head and pelvis CBCT data derived from the ARCP as used in Section 4.4. We derived treatment plans considering neck and uterus cancer respectively based on specifications given in [15]. For the different radiation types, we changed the geometry to induce diversity and represent the restricted movement in each case: the photon delivery operated around 360° as with the Varian TrueBeam; the proton has a reduced range in motion to represent the practical limitations of this technology; and carbon ion has only a single angle to investigate the particle's effectiveness from this fixed position. Specifics of the dose planning for each case are given as follows:

Head planning procedure

- A planning target volume (PTV) of was defined in the neck as a sphere of radius 0.9 cm over which a dose was specified of 50 Gy delivered in 40 fractions.
- Organs at risk (OAR) were defined as the spinal cord, lymph nodes, and the eyes.

- For the photon plan, the beams were defined at 10 equispaced gantry angels around the specimen, to approximate an intensity modulated radiotherapy (IMRT).
- For the proton plan, the beams were defined at 5 gantry angles in the upper semicircle, to account for the reduced rotation of the head.
- For the carbon plan, the beam was at a single position above the specimen.
- For each of photons, protons and carbons, treatment plans were optimised on the oracle volumes to deliver the prescribed dose to the target, whilst minimising dose delivered to the OAR, using the built in interior point optimiser (IPOPT). Resulting plans from each of the three particles are shown in Figure 5.2 with the dose colour reference in Figure 5.1.

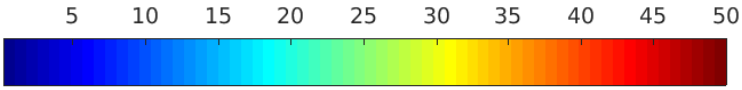


Figure 5.1: Colour scale for dose maps (Gy).

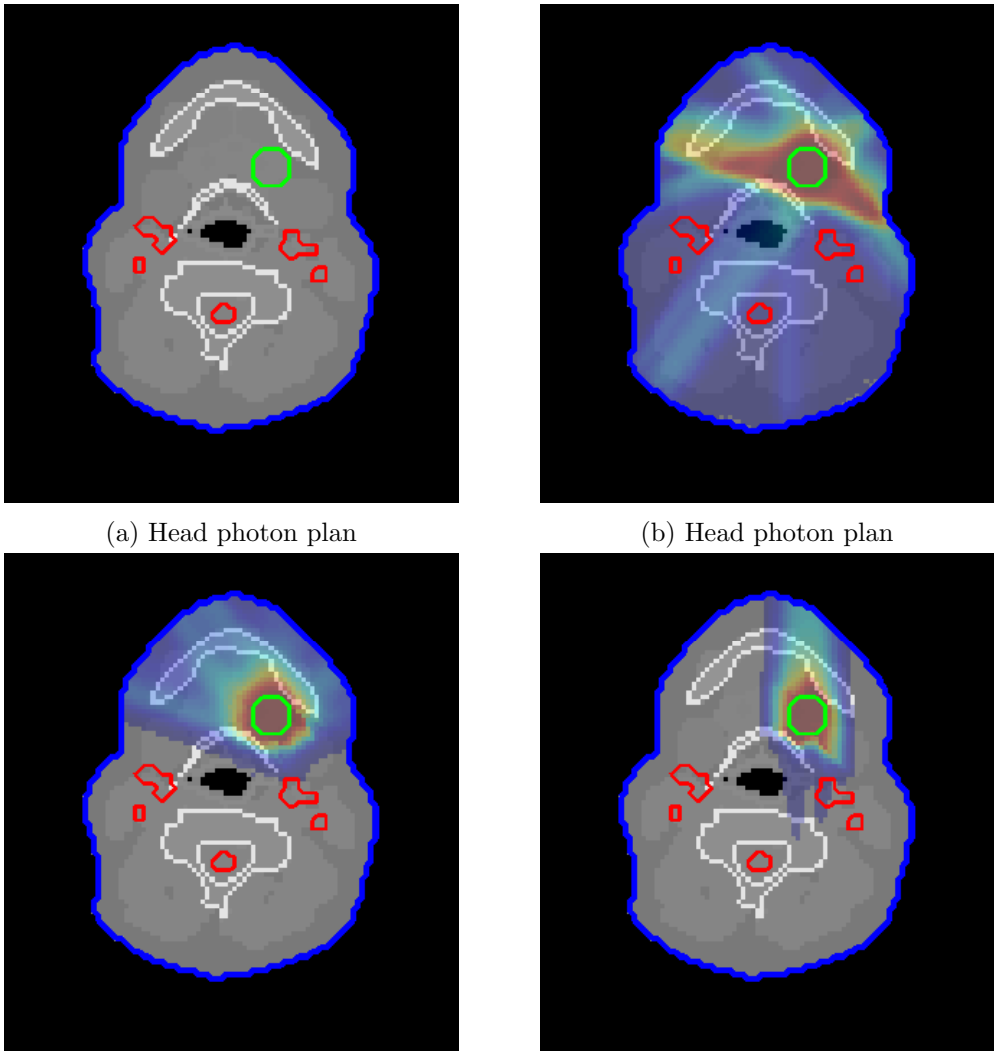
Pelvis planning procedure

- A PTV was defined as the uterus organ plus a 2 mm boundary over which a dose was specified of 50 Gy (J/kg) delivered over 40 fractions.
- Organs at risk (OAR) were defined as the femoral heads, the rectum, the bladder, and the small intestine.
- Each of the three delivery geometries were identical to the head case.
- The three plans were similarly optimised using IPOPT. Resulting plans from each of the three particles are shown in Figure 5.3 with the dose colour reference in Figure 5.1.

The following reconstruction methods were evaluated in every case:

Reconstruction methods under test

- *Slab*: all tissues within the body region defined by the blue contours in Figure 5.2 and Figure 5.3.



(a) Head photon plan

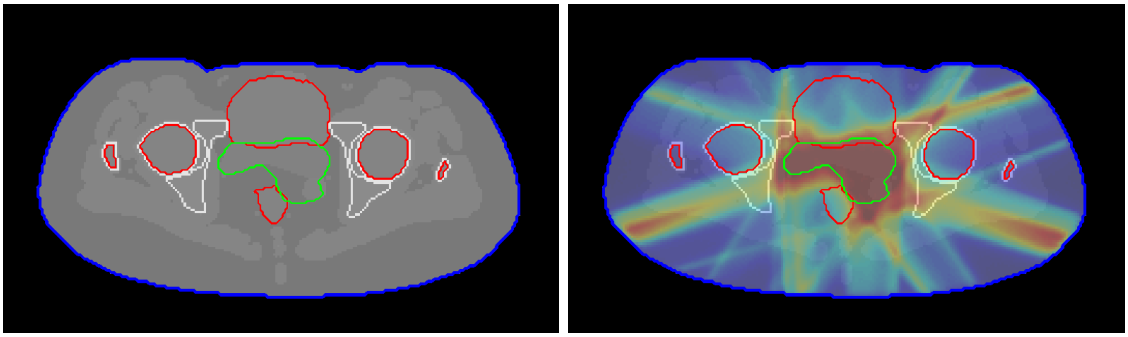
(b) Head photon plan

(c) Head proton plan

(d) Head carbon plan

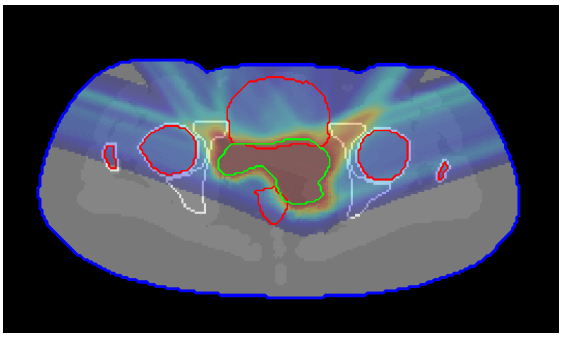
Figure 5.2: Dose plans for the head data for photons, protons and carbon ions. The PTV is shown in the green contour, the OAR in red, and the body in blue. The dose has a transparency threshold of 1 Gy, and the colour map is according to Figure 5.1.

- *FDK*: FDK reconstruction of the data after water LUT beam hardening correction and fASKS scatter correction, with a Hann window to reduce noise from photon starvation.
- *PWLS*: PWLS iterative reconstruction applied to the same linearised data as FDK, with TV regularisation.
- *IMPACT*: the IMPACT applied to the raw synthetic measurements, supplemented by fASKS scatter estimate from preprocessing, and the same TV regularisation as PWLS.
- *Polyquant*: Polyquant reconstruction technique applied as IMPACT to the raw data,

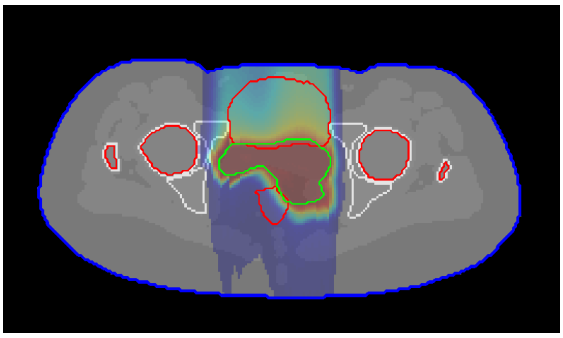


(a) Pelvis photon plan

(b) Pelvis photon plan



(c) Pelvis proton plan



(d) Pelvis carbon plan

Figure 5.3: Dose plans for the pelvis data for photons, protons and carbon ions. The PTV is shown in the green contour, the OAR in red, and the body in blue. The dose has a transparency threshold of 1 Gy, and the colour map is according to Figure 5.1.

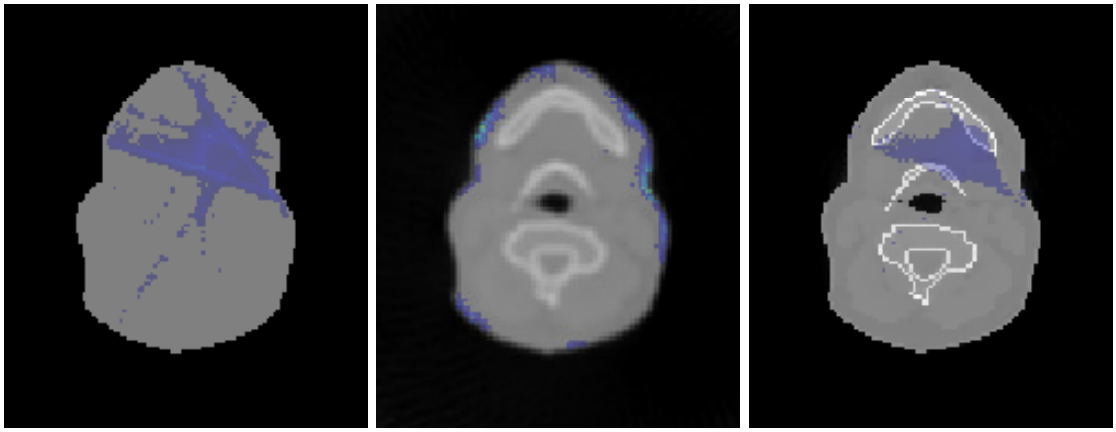
and with the pre calculated scatter estimate as in Chapter 3. These are the same volumes as the 'Pre-fASKS' presented in Section 4.4.

- *PolySKS*: the Polyquant reconstruction technique with polyenergetic scatter model fusion as in Chapter 4. These are the same volumes as the 'Poly-SKS' presented in Section 4.4.

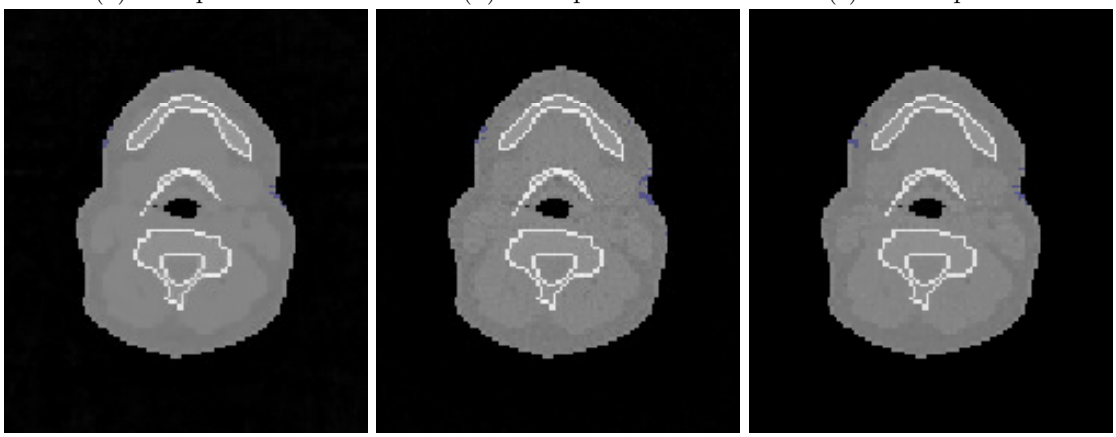
Each of the iterative methods utilised the same TV regularisation, though with the parameter $\lambda = 1 \times 10^{-5}$ scaled by $2\times$ in the case of PWLS and IMPACT to fairly compensate for the differing relative regularisation of these approaches as noted in Figure 3.9. The algorithms in each case were as the F-P 2 in Algorithm 3, with NLL objective functions, and were run for 10 and 20 epochs for the head and pelvis experiments respectively.

5.2.1 Photon Therapy Analysis

The various images under test with superimposed dose error for the head photon planning are shown in Figure 5.4, with the pelvis examples shown in Figure 5.5.



(a) Slab photon (b) FDK photon (c) PWLS photon



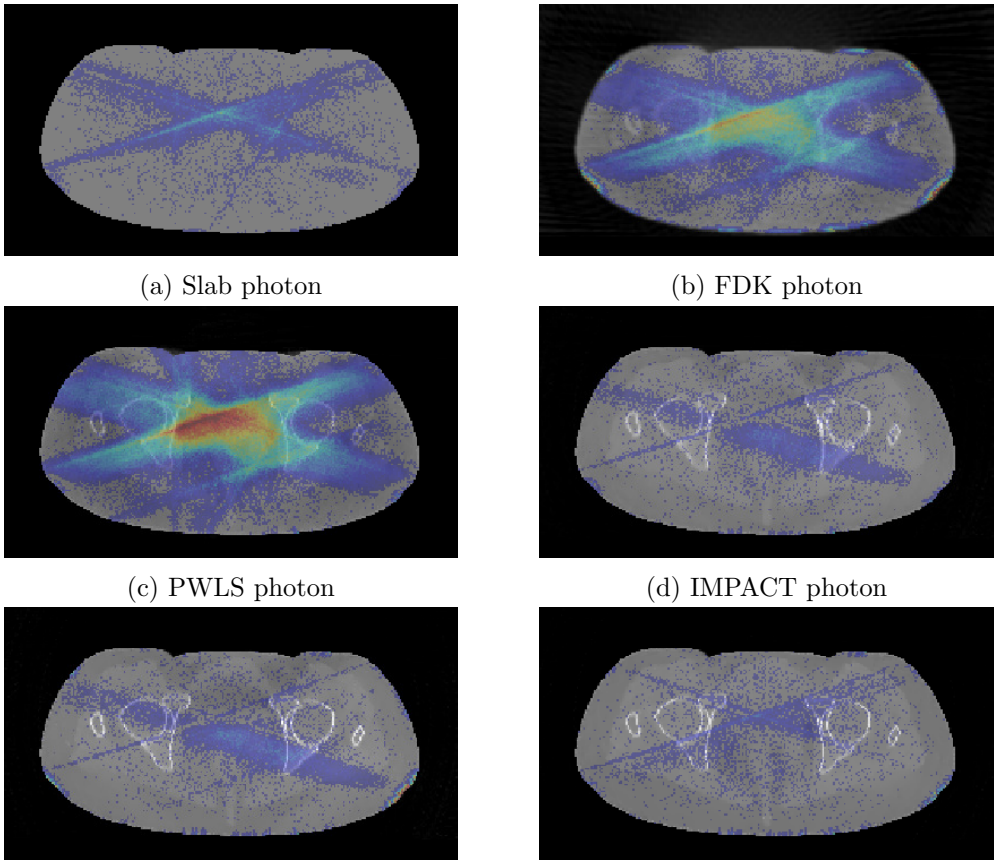
(d) IMPACT photon (e) Polyquant photon (f) PolySKS photon

Figure 5.4: Visual results from the head photon dose test. The dose errors are superimposed with the same colour reference as in Figure 5.1 but with range of $[0.5\%, 10\%]$ and transparency below this.

From the photon head examples shown in Figure 5.4, it can seen that IMPACT, Polyquant and PolySKS all exhibit very low error, with only few pixels having errors greater than 0.5%. On the other hand, the slab and PWLS lead to significant inaccuracies at the centre, and FDK has errors at its edges.

In the photon pelvis case shown in Figure 5.5, the errors are clearly larger than the head example, arising from the larger object mass. Of the methods tested, PolySKS has the highest consistency, with the FDK and PWLS showing large errors, especially at the centre.

The quantitative photon plan results are shown in Table 5.1. It can be seen that PolySKS gives rise to significantly lower errors throughout the PTV than any other method under test, which is an order of magnitude lower than that of PWLS. The OAR errors are generally much lower than with the PTV, though this is due to the fact, by design, that



(a) Slab photon (b) FDK photon

(c) PWLS photon (d) IMPACT photon

(e) Polyquant photon (f) PolySKS photon

Figure 5.5: Visual results of the pelvis photon dose test, with same error display as Figure 5.4.

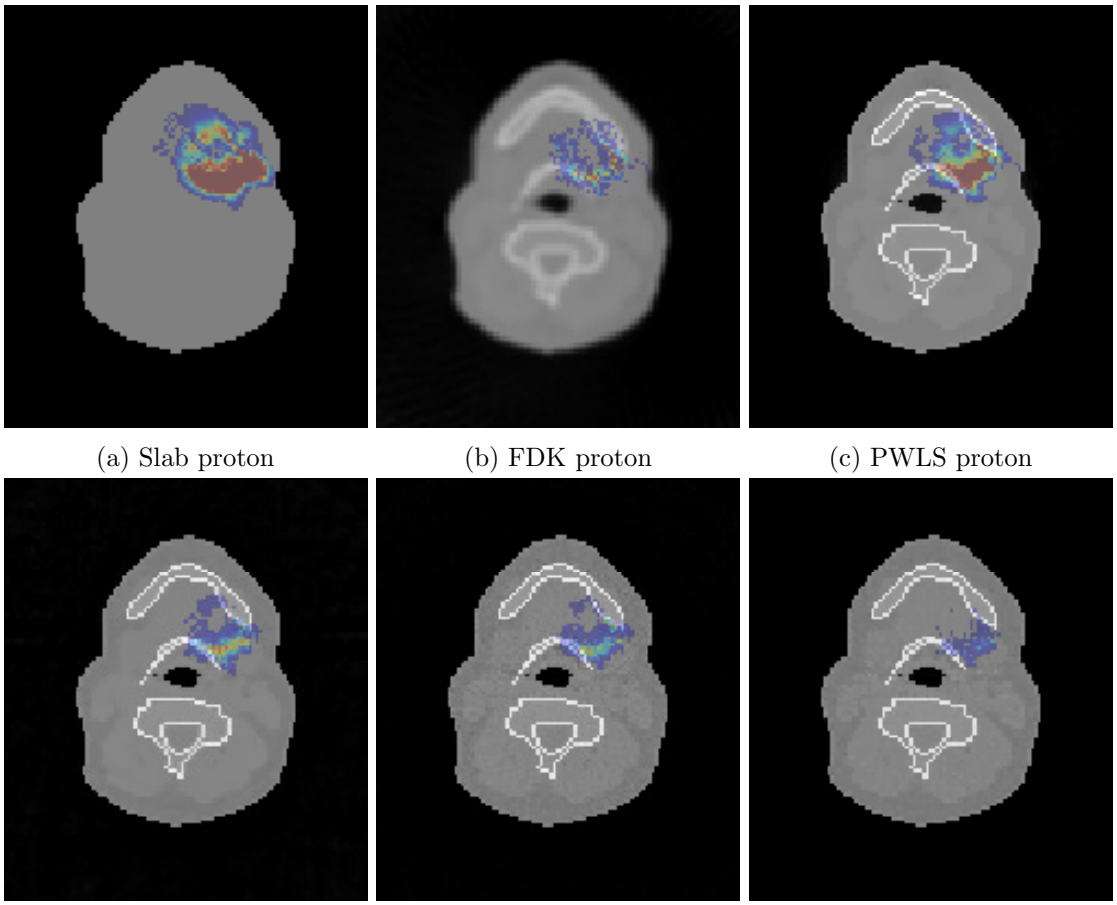
Table 5.1: Quantitative results: photon planning test (RMSE)

	head			pelvis		
Scheme	body	PTV	OAR	body	PTV	OAR
Slab	0.0839	0.632	0.0423	0.201	0.841	0.277
FDK	0.225	0.574	0.0697	0.495	2.46	0.638
PWLS	0.104	0.452	0.0557	0.659	3.57	0.967
IMPACT	0.126	0.125	0.0563	0.157	0.645	0.224
Polyquant	**0.0551**	0.0762	0.0426	0.201	0.733	0.237
PolySKS	0.0565	**0.0301**	**0.0418**	**0.139**	**0.398**	**0.190**

minimal radiation is directed through these regions. Another interesting observation from these results is the slab method, which is by far the simplest method under test, performs reasonably well. In the pelvis case, the slab errors are significantly lower than both FDK and PWLS, and of a similar level even to Polyquant. This suggests that for photons

there is not a huge sensitivity to fine electron density accuracy, and bulk assignments, a technique common in MR based planning [167], can be reasonably accurate in some cases.

5.2.2 Proton Therapy Analysis



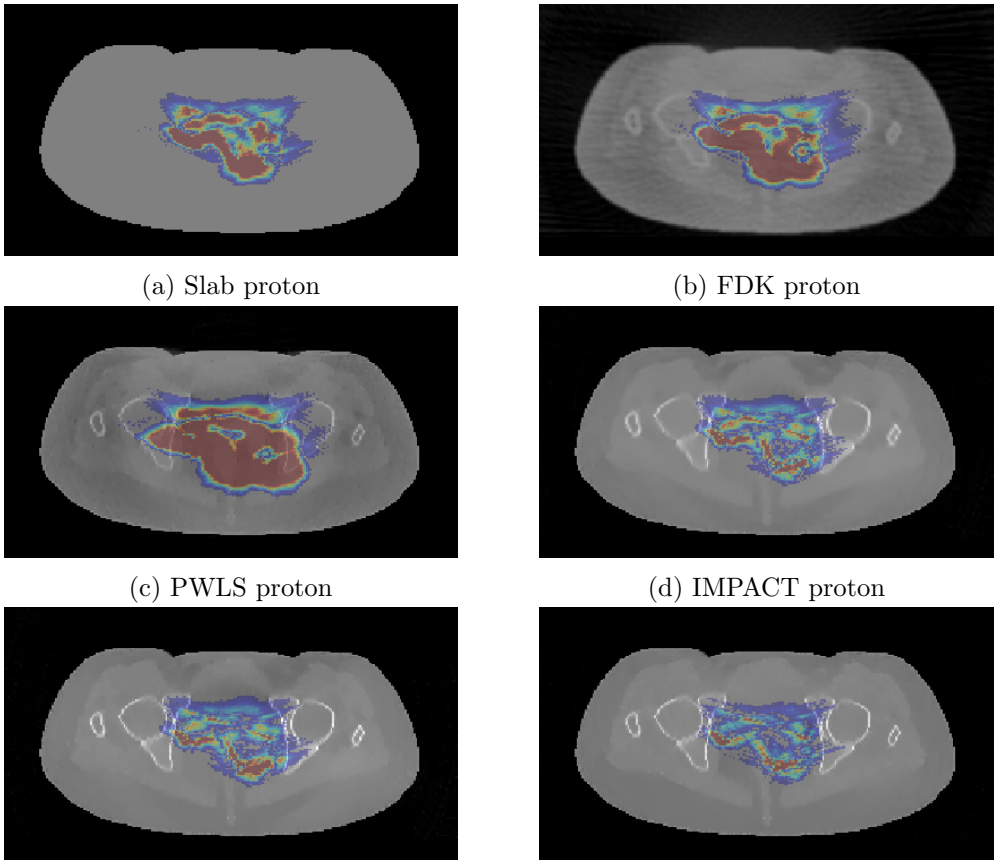
(a) Slab proton (b) FDK proton (c) PWLS proton

(d) IMPACT proton (e) Polyquant proton (f) PolySKS proton

Figure 5.6: Visual results of the head proton dose test, with same error display as Figure 5.4.

The visual results from the proton plans are shown in Figure 5.6 and Figure 5.7 for the head and pelvis respectively. From the head examples, there are generally higher levels of error than were present in the photon plans, especially in the polyenergetic reconstruction methods IMPACT, Polyquant and PolySKS. Additionally, the slab error is quite significant, with large regions having errors over 10% in the centre. Similarly to the photon plans, the PWLS pelvis has the highest error visible in Figure 5.7c.

The quantitative results from the proton plans shown in Table 5.2. A similar trend as the photon planning can be seen with this result, with the PolySKS showing a huge gain



(a) Slab proton (b) FDK proton

(c) PWLS proton (d) IMPACT proton

(e) Polyquant proton (f) PolySKS proton

Figure 5.7: Visual results of the pelvis proton dose test, with same error display as Figure 5.4.

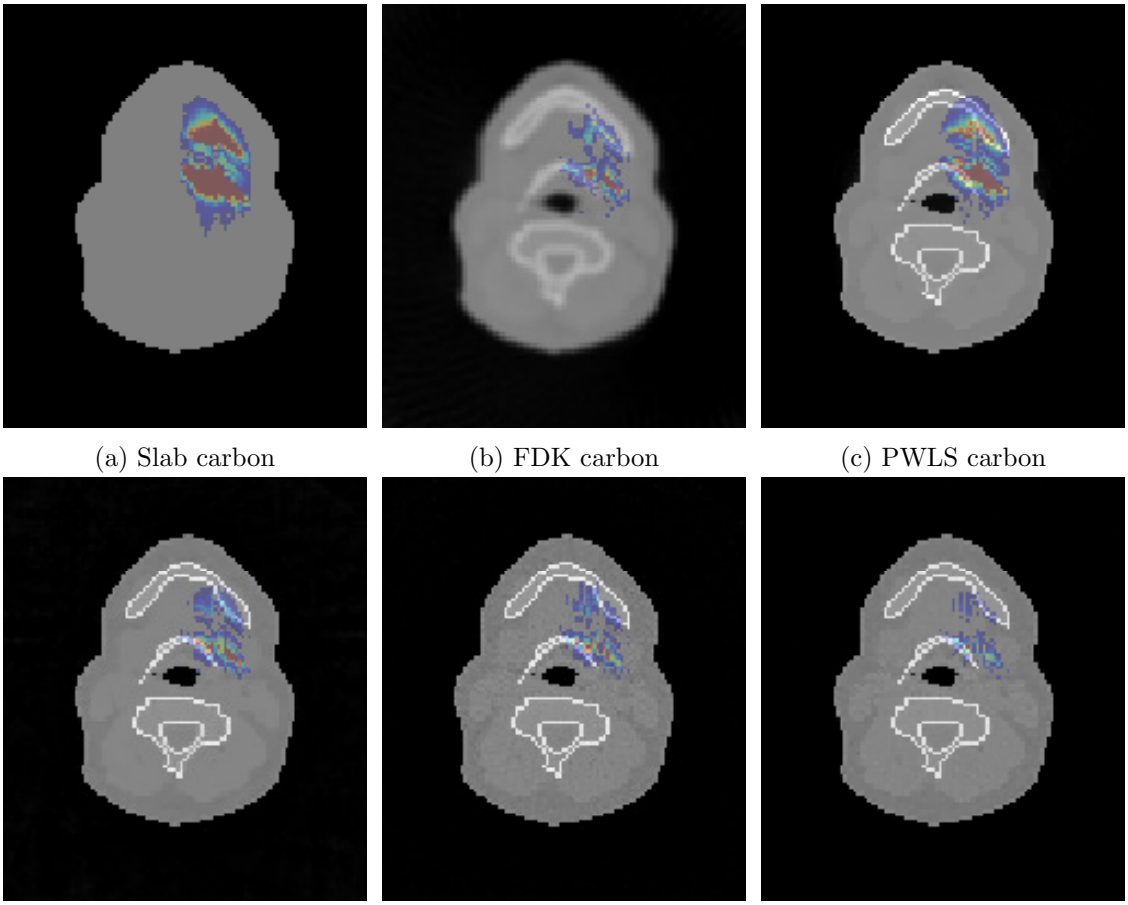
Table 5.2: Quantitative results: proton planning test (RMSE)

	head			pelvis		
Scheme	body	PTV	OAR	body	PTV	OAR
Slab	0.635	2.00	0.0594	0.780	3.50	1.68
FDK	0.509	1.84	0.112	1.15	4.63	2.80
PWLS	0.320	2.49	0.0429	1.94	6.62	5.02
IMPACT	0.178	0.695	0.0139	0.402	2.47	0.754
Polyquant	0.115	0.404	0.00794	0.381	2.43	0.675
PolySKS	**0.0909**	**0.177**	**0.00724**	**0.285**	**1.55**	**0.556**

over every other method in the PTV regions. Also similar is the very poor performance of PWLS, which is considerably worse than the slab assignment. One difference however, is that the slab accuracy is no longer comparable to the polyenergetic methods, which indicates that proton planning is far more sensitive to having accurate electron density

estimates with structure.

5.2.3 Carbon Ion Therapy Analysis



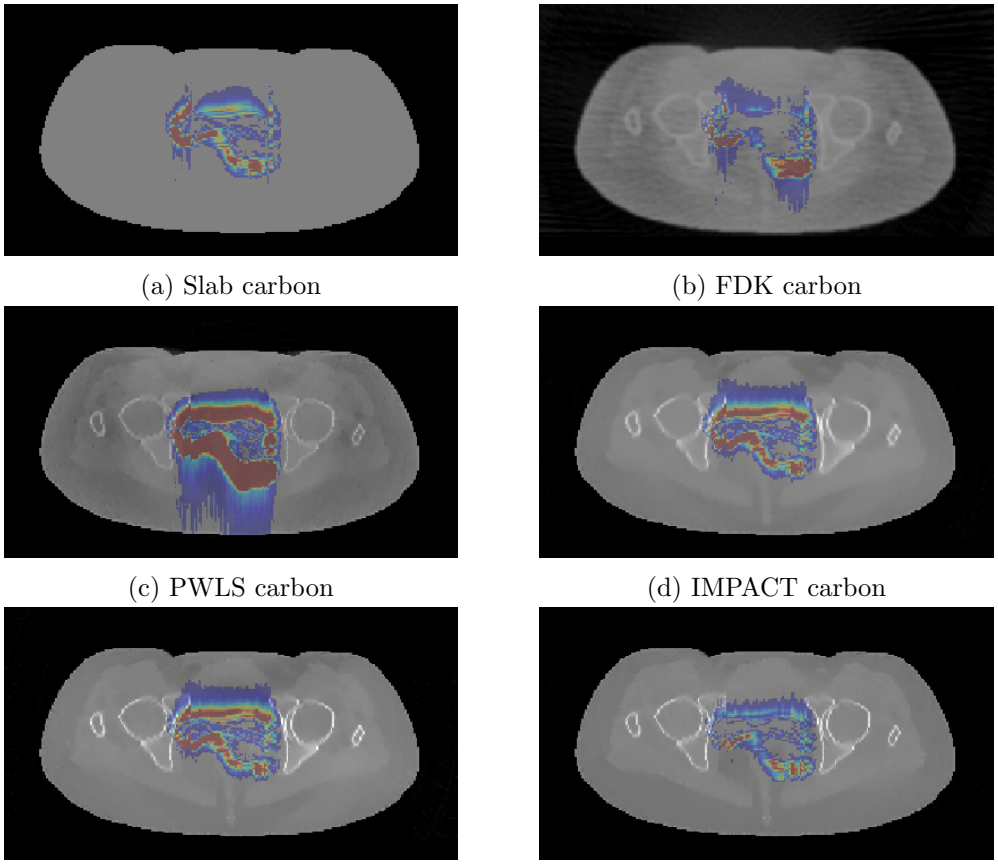
(a) Slab carbon (b) FDK carbon (c) PWLS carbon

(d) IMPACT carbon (e) Polyquant carbon (f) PolySKS carbon

Figure 5.8: Visual results of the head carbon ion dose test, with same error display as Figure 5.4.

Visual results from the carbon ion experiment are shown in Figure 5.8 and Figure 5.9 for the head and pelvis respectively. Visually, the errors are similar in extent and trend to the proton plans, with PolySKS showing the least error, and the slab and PWLS having the worst performance in the two cases. One difference between both the photon and proton plans however is that the FDK exhibits a lower level of error.

From the quantitative carbon ion results in Table 5.3, a familiar trend can be seen, with the PolySKS having the best performance by some margin. An interesting difference here however, is the enhanced performance of the FDK, especially in the pelvis case, where it is the second best method under test. This is likely due to the FDK exhibiting fewer



(a) Slab carbon (b) FDK carbon

(c) PWLS carbon (d) IMPACT carbon

(e) Polyquant carbon (f) PolySKS carbon

Figure 5.9: Visual results of the pelvis carbon ion dose test, with same error display as Figure 5.4.

Table 5.3: Quantitative results: carbon ion planning test (RMSE)

	head			pelvis		
Scheme	body	PTV	OAR	body	PTV	OAR
Slab	1.06	3.37	0.316	0.979	1.84	2.24
FDK	1.13	2.69	0.648	0.698	1.21	1.13
PWLS	0.629	3.20	0.277	2.17	4.68	5.50
IMPACT	0.427	0.789	0.134	0.755	2.68	1.94
Polyquant	0.296	0.828	0.0908	0.710	2.56	1.84
PolySKS	**0.223**	**0.360**	**0.025**	**0.328**	**1.01**	**0.734**

shading artefacts in the central region, which is the only covered with our fixed gantry simulation.

5.3 Physical Validation on Phantom

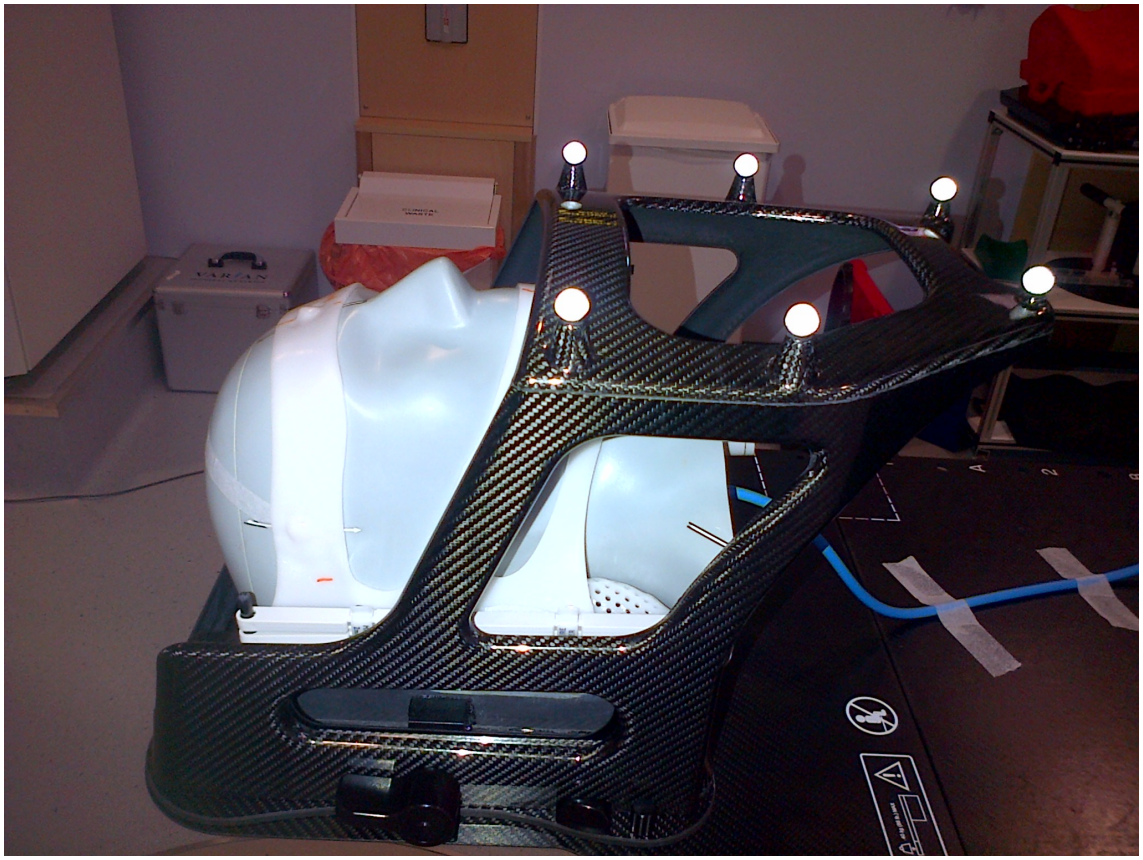


Figure 5.10: Photograph of experimental set up with STEEV head phantom The blue cable emerging from the base of the neck is the from the ionisation chamber.

In this experiment, we wish to evaluate the accuracy of dose calculation against physical measurements and real CT data. This was conducted on the STEEV head phantom, which we irradiated from 9 gantry angles with a 10×10 cm uniform field into its isocentre on a Varian TrueBeam linear accelerator. Located at its isocentre was a PTW PinPoint ionisation chamber, that produced a ground truth energy absorption measurement for each gantry angle. For the evaluation, we compared the measured dose against doses calculated using the Varian Eclipse treatment planning software, on a number of different images, with beams matching those used in the experiment.

Methods for comparison were a 'planning CT' from scanning the phantom on a Philips Brilliance Big Bore 'fan-beam' CT system, which is used as a basis for treatment planning. In this case, the reconstructions were generated from the system's software and technique. CBCT scans using the on board imaging of the TrueBeam system were also acquired, from which we also took the system's standard reconstruction result that we denote as 'FDK

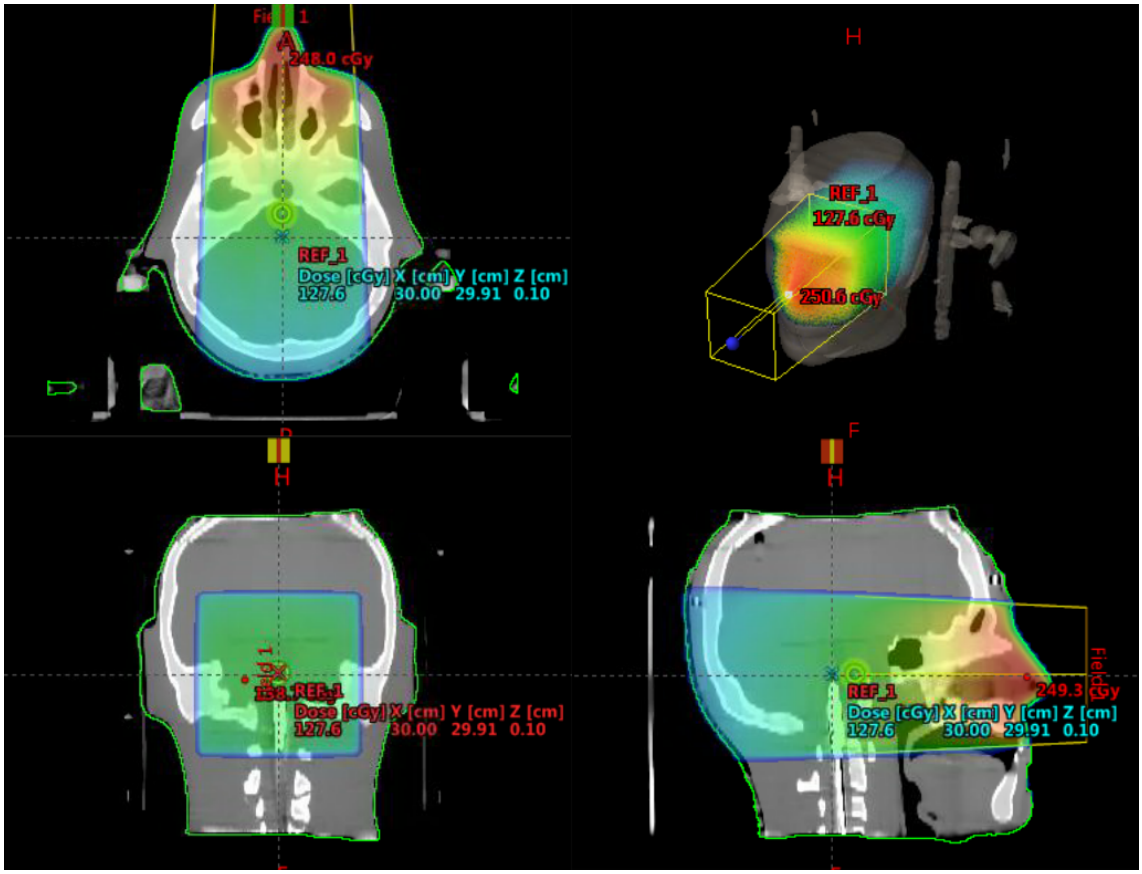


Figure 5.11: Screen captures from the Varian Eclipse dose calculation software, showing the experimental beam arrangement for a gantry angle at 0°.

CBCT'. Finally, we performed our own quantitative reconstruction of the data using the Polyquant method as outlined in Chapter 3. In fact, the image used was the very same as that presented in Figure 3.16.

For the evaluation, we isolated the region from the centre of the location of the ionisation chamber in each image, took the average of the calculated dose from Eclipse in this area, then converted to the same units as the physical measurements. These are then shown in Table 5.4.

Quantitative results from the dose planning accuracy are shown in Table 5.4, from which a remarkable observation can be made. Not only is the error reduced by several multiples from the FDK CBCT result, but the Polyquant reconstruction significantly outperforms the planning CT accuracy with a 26.5% reduction in error. This may be a surprising result due to the fan-beam CT's usual superiority, from its heavily reduced scatter over CBCT, and the routine use of this modality in practice for planning [15]. We would expect that if we had access to the raw data from the fan-beam CT also, the Polyquant applied to this

Table 5.4: Quantitative dose results from physical measurements on STEEV phantom.

Gantry angle (deg)	dose (cGy)	Planning CT	FDK CBCT	Polyquant CBCT
270	155.84	154.0	148.2	154.2
300	159.78	157.5	150.6	158.3
330	137.18	136.7	130.4	138.5
0	127.98	131.5	124.4	130.9
30	137.98	136.3	129.9	138.3
60	160.31	157.1	151.0	158.5
90	155.58	154.8	149.3	155.0
RMSE	–	2.234	7.491	**1.642**

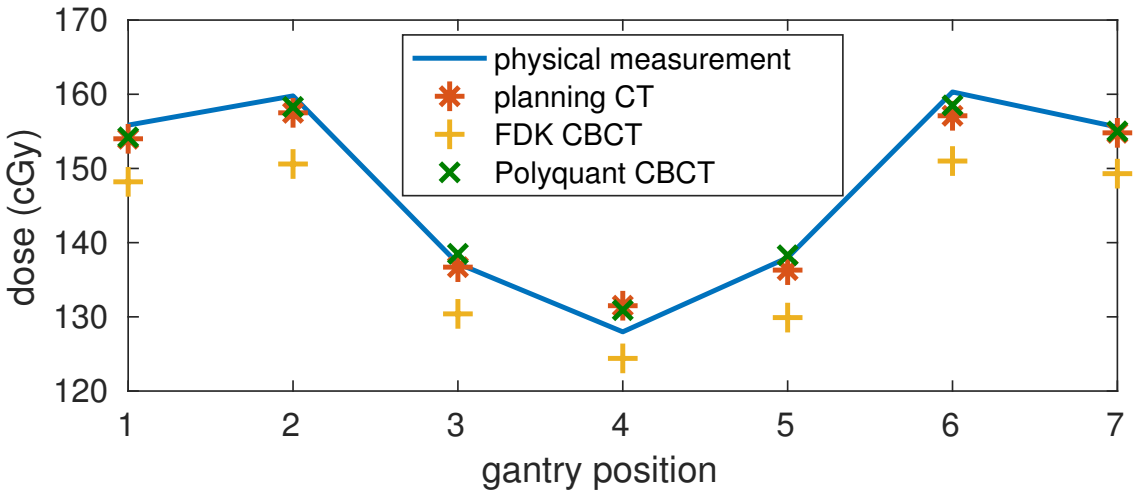


Figure 5.12: Graphical representation of quantitative STEEV dose results as reported in Table 5.4.

would likely result in better accuracy still.

5.4 Conclusions and Discussion

In this chapter we have made various evaluations of the dose calculation accuracy with our reconstruction techniques, which has exposed several interesting areas for discussion.

Firstly, the real data analysis shows that a huge increase in CBCT dose calculation accuracy is possible with use of our Polyquant method. We do note however, as indicated from the photon accuracy numerical results, that this gap is likely to be considerably narrowed by using slab assignment of the images to water. However, we also expect a

further gain of Polyquant if evaluated with the integrated scatter estimate PolySKS, as this leads to a significant quantitative gain for the STEEV data in Table 4.6. Usually it is unwise to put too much weight on a single experiment, so replication on a range of different phantom types and sizes, and with a charged particle such as proton therapy, would be interesting to evaluate its complete performance. In any case, this is a very promising result, and motivates the pursuit of quantitative CBCT reconstruction and replanning, especially since non-rigid deformations between the planning CT acquisition of a real patient will degrade its accuracy even further.

With the numerical experiments, we were able to perform an extensive evaluation with many reconstruction techniques, different specimens, and various radiation types also. This allows us to form clear pictures of the effectiveness of the various scenarios. The first key finding from this is that photon dose calculation is relatively insensitive to fine structure variations, and the bulk assignment to water outperformed both FDK and PWLS. For particle therapy of carbon and proton, the slab accuracy is comparable to these techniques, but the sensitivity to structure is clearly greater with these types of radiation.

A second common trend throughout all of the numerical experiments is the polyenergetic reconstruction offers a large gain in dose calculation accuracy. Moreover, our PolySKS was almost unanimously the best performing method throughout. In the proton and carbon ion experiments, the polyenergetic techniques IMPACT, Polyquant and PolySKS resulted in an order of magnitude reduction in error over the other methods under test, which suggests that these techniques may offer large gains in the efficacy and reduction of toxicity from these approaches.

A powerful connection to make would be to tie the various dose calculation accuracies to radio-biological effects such as those in the QUANTEC summary. Although there are not many cases where we could find tight links, we may summarise some of the indications. With the numerical pelvis data the slab model, FDK and PWLS showed significant errors within the bladder, bowel and rectum. As documented by Marks et al. [166], levels of toxicity is linked to an increase in dose deposited in these organs. One could therefore predict that through having accurate doses into these regions as in Polyquant–PolySKS, would lead to a reduce in complications in these organs.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

The ultimate aim of this work was to close the technological gap between CBCT reconstruction, and the ability to provide quantitative imaging for radiotherapy dose calculation in planning. From the extensive evaluations for this setting in Chapter 5, we believe we have demonstrated this.

Using the Polyquant method we developed, we have demonstrated that one can achieve a higher accuracy through an empirical fitting to either electron and mass density or indeed monoenergetic attenuation, than by employing a physically inspired attenuation model such as photoelectric–Compton or water–bone. This fact not only allows one to perform direct inference of electron density at a high accuracy, but also to mitigate artefacts from spectral variation and metal objects. Although this comes at several times the computational cost over gradient descent on a classical objective function such as PWLS, the factor in accuracy gain can be comparable to this.

Two key conclusions were drawn from the scatter fusion work presented in Chapter 4: that one can expect a significant gain from integrating scatter estimation methods into iterative reconstruction; and the polyenergetic and electron density information available during Polyquant allow better scatter artefact mitigation than competing convolutional models. Through use of the Polyquant–PolySKS method, we demonstrated very high quantitative electron density imaging accuracy with real data, which was even in some cases within the manufacturing tolerance of the material. We also suggested several extensions to this approach that may offer better practical performance still.

Finally, the dose calculation analysis in Chapter 6 presented several interesting conclusions. Firstly, that photon therapy planning was not overly sensitive to image structure, and a simple slab model outperformed traditional reconstruction methods such as FDK and

PWLS in these cases. On the other hand, particles such as protons or carbon ions are highly sensitive to image structure, where the combined Polyquant–PolySKS method provided a huge advantage by at least several multiples over any other method testing. Moreover, with a quantitative gain over PWLS as high as 9× in the proton pelvis OAR, we believe this more than justifies for the 2.5× higher computational cost of this approach. The real data then presented a most interesting result: that Polyquant provided a 78% reduction in dose error over the commercial reconstruction and calibration, and a 26% reduction in error over an enclosed gantry fan-beam planning CT image. This is an exceptionally promising result, as not only would this imply we are successful in matching the accuracy used for clinical planning, but that we could even significantly out perform it. By matching the quality of planning CT, one can begin utilising the CBCT geometry for recalculating and refining the treatment. This also suggests that the Polyquant method should offer considerable gains to the planning CT system itself, which has fundamentally less scatter, and should be of a higher accuracy.

With these observations, the methodologies are very promising for enhancing the accuracy of planning, offer better adaptability to organ structure, which could lead to reduced toxicity to healthy tissues and the ability to treat more patients with the same resources through fewer treatment fractions. We feel there are also many promising extensions to this work, and potential applications of the methods, which we will discuss in Section 6.2. We also readily admit there remains a practical gap between these techniques being used in clinical practice, and we will also be discussing this in Section 6.3.

6.2 Future Works

6.2.1 Prior CT Exploitation

During the course of treatment, there are usually several images acquired of the same patient. This may often include a planning CT from an enclosed gantry fan-beam CT, and several CBCTs from the times of delivery. By considering all these images together, one should expect some redundancy in information between them, which may be exploited to either enhance the quality of subsequent scans or allow them to be taken at a substantially lower dose. In Section 2.5.3, we introduced some prior image regularisation methods that attempt to reduce the sampling requirements; one can also use previous images to correct

for some of the artefacts such as scatter or beam hardening. Since we have explored these ideas in some detail, we will present some of the findings here, and indicate how they may be extended further with the work in this thesis.

To begin with, the prior regularisation methods such as PICCS [12] or PIRPLE [107,108] described in Section 2.5.3 may be readily included with the Polyquant and Polyquant–PolySKS methods. Since the algorithms presented in Algorithm 1, Algorithm 2 and Algorithm 3 accept any convex regularisation function in the proximal term, this may adopt an appropriate function such as the TV against a prior image. There are a couple of points worth raising here though. Firstly, as the images themselves are in electron density, it is important that any prior is also quantified as such. For use of a planning CT for example, one would either have to use this after calibration to electron density, or indeed from applying a Polyquant reconstruction to this data also.

A second consideration when including prior regularisation is that using more than one regularisation term as in PICCS is not entirely trivial, if more than one of them is a non-smooth function like TV. One option is to use smooth approximations of TV, as the authors employ in the PIRPLE article [107]. With this, the gradient of the regularisation may be added to the NLL derivative, and the proximal step will be replaced by a projection onto the constraint set $\mathcal{C}$. Another possibility is to employ the generalised forward-backward splitting algorithm of [168], which would allow the use of more than one TV function. This method works by computing the proximal step w.r.t. each regularization function separately, then averaging each of these updates.

From preliminary experiments of prior regularisation from repeat scans of the bladder, we found that it did not offer a huge gain in performance over straight spatial TV regularisation. Another finding is that it is very sensitive to the registration of the prior or lack thereof, and poor alignment can in fact lead to large artefacts in the reconstruction. This does motivate the use of PIRPLE methods, which is likely a promising approach to adopt. One drawback of this is that including many sophisticated registrations throughout reconstruction, will very likely vastly increase the total computational time.

Another usage of prior scans is using the volumetric information to get a good estimate of the scatter as in [21,163,169]. We have performed a preliminary survey of these approaches in [158] on numerical data, but not with Polyquant methods. Findings from this were that they are also very sensitive to the registration accuracy, and performed no better than

pre-calculated SKS. In light of our findings with PolySKS, we are doubtful these methods would perform any better than this, but they may have potential in very challenging scattering scenarios. Another factor in these methods is their large computational load. In the article in [158] we do propose an off-line estimation of scatter, followed by on-line matching to the current scan, but this approach would require validation with realistic data from repeat scans that was not available to us at the time.

Finally, another interesting concept is in using previous imaging information to account for the polyenergetic effects of the source. Specifically, to use structures and spectral information from previous scans to reduce beam hardening artefacts. One example of this is known component imaging [170], where the authors show how the effects of a metal object such as a screw can be compensated for in the measurements, given that its structure is known. We have also experimented with a specific example of this concept we denoted 'bone borrowing'. Here, the hard structures such as bone or metal are taken from previous high quality images such as planning CT, aligned to the current pose of the patient, and used to compensate the measurements. With this, one only needs to reconstruct the soft tissue structures such as muscle and fat, which are reasonably modelled as water as in the Poly-SIR model. Whilst experimenting with this approach however, we found that its performance fell below Polyquant in a numerical setting, by both its reliance on accurate registration of the bone, and the superior attenuation modelling of soft tissues in Polyquant over a water model.

6.2.2 Dynamic Imaging

Throughout this work, we have treated the imaging of static specimens, but in practice there may be significant dynamic movement of the patient that may be compensated for. A particularly compelling example of this is the imaging of the chest, where there will be movement from both the respiration cycle and the cardiac cycle. There are physical approaches to mitigate this movement, including having the patient hold their breath or only acquire measurements during a given phase, known as gating, but these either present challenges in consistency or prolong the acquisition significantly. Another approach we will discuss here is by considering an extra temporal dimension on the reconstructed volume, known as 4DCT, and how the ideas from this thesis may be extended into this setting.

4DCT involves adding a temporal dimension to the data, which can be achieved

by separating the projections into phases of the motion that can be individually reconstructed. This separation can be made based on an auxiliary measurement, such as the electrocardiogram for the cardiac cycle, or optically tracking an object placed on the chest for the respiration cycle, on which phases can be designated based on phase or amplitude intervals [171]. It is argued that for CBCT, it is preferable to perform this retrospectively due to the slow rotation of the gantry [79]. Naively reconstructing each phase separately carries a huge penalty, which is that the effective volumetric sampling for each 3D phase is reduced by a factor of the number of phases. With this, the streak artefacts that become probable may well lead to far lower image quality than with the uncorrected blurring, unless the dose is significantly increased.

Needless to say, both the quantitative model Polyquant, and integrated scatter model Polyquant–PolySKS are entirely compatible with gated 4D projection data, and it is likely they could offer considerable accuracy gains over comparable iterative methods. In exploring these ideas, a remaining challenge would be forming a reasonable dynamic model that is able to regularise throughout the phases and reduce the inherent sampling deficit. Some authors have suggested using temporal TV to achieve this [79] [106], and it has been shown that simple spatio–temporal regularisation works well [80], which would also be compatible with our presented algorithms. However, one would not expect the difference between neighbouring phases to be piecewise constant for complex motion, as implied by these model, but to be somewhat sparse and have some periodic structure over the whole cycle. Therefore, we feel there may be more appropriate regularisation choices that could be explored to exploit this.

6.2.3 Spectral Imaging

We have developed methods for the case where one has a polyenergetic source and an energy indiscriminant detector, as is the most common clinical technology. There do now exist both spectrally discriminant detectors and systems with multiple sources such as dual energy CT (DECT), and we will discuss the implications of these on our methods, and whether they may still be applicable.

To begin with, if one had a truly spectrally discriminant detector, then one would have a wealth of information that has the potential to make the imaging more powerful. With a spectrum at each detector, the effects of beam hardening should be able to be modelled.

With this, our developed techniques may not be entirely necessary. An interesting question would be whether the extra monetary cost of employing these detectors is justified, when the quantitative performance is compared against our approaches with cheaper hardware.

The motivation behind using two sources in DECT is to gain extra spectral information from diverse measurements. Since the shape of X-ray spectra as shown in Figure 2.4 are similar for different tube potentials, diversity is often achieved by applying metal filters such as tin on a higher energy source [146]. From the model in [42], DECT does appear to offer the ability to perform quantitative imaging in electron density as with Polyquant by using the photoelectric–Compton decomposition in (2.6) directly. The reality is that it is difficult from an engineering perspective to have two perfectly matched projections from different sources [146], the solutions of the simultaneous equations in [42] are unstable with noise, and we have demonstrated for biological tissues that the Polyquant fitting is superior for biological tissues than this photoelectric–Compton approximation. With these factors, it is unsurprising that from preliminary experiments, Polyquant substantially outperformed a DECT projection decomposition approach, though a statistical iterative method such as [172] may reduce this gap.

Although for biological tissues, the effectiveness of traditional DECT over Polyquant remains to be seen, there is certainly potential for the latter to be enhanced by spectrally diverse measurements. A key observation from Figure 2.8a is that a range of materials do occupy the 2D space of electron density against effective atomic number. This implies that a single source technique such as Polyquant cannot unambiguously separate all materials, as was demonstrated with plastic materials in Figure 3.5. It may be possible to overcome this by finding some relationship, ideally piecewise linear, between the effective atomic number and polyenergetic attenuations. With this, one could extend Polyquant to simultaneously reconstruct both electron density and effective atomic number, given at least two spectrally diverse measurements, which may allow the characterisation of many different materials.

6.2.4 Beyond the CBCT Geometry

The focus has been on enhancing the accuracy of CBCT, but we believe these ideas are also applicable beyond this geometry. We will discuss how they may offer benefit to other X-ray modalities such as enclosed gantry fan-beam CT, and also where the physical models

may even be used beyond CT imaging.

In the first case, the Polyquant model is directly applicable to all generations and forms of X-ray transmission tomography. This certainly includes enclosed gantry fan-beam CT, and this geometry was directly modelled in Section 3.4.2. We wholeheartedly believe this may be adopted by planning CT systems, for instance, and expect this to offer a sizeable gain over conventional linearised reconstruction followed by nonlinear calibration, as is the standard. With the scatter model, PolySKS, this should also be applicable with other geometries. The polyenergetic forward scatter factor is certainly universal, though the Gaussian kernels may need to be adjusted with curved or some offset detector geometries.

6.3 Towards Clinical Use

The ultimate goal of this work is to enhance the efficacy of radiation therapy, through enhanced imaging. Although we have demonstrated these ideas with real data in each case, there is still a gap between their use in clinical practice. We believe a few of the remaining hurdles are: having a robust software implementation, the computational speed, and proof of concept in a clinical trial. We will try to address these points and suggest how the gap may be closed.

The first two goals of having a robust software implementation and a low computational speed could go hand–in–hand. It would involve a rapid low level version of the algorithms written in an embedded language, and ideally exploiting fast parallel hardware such as GPGPU, with a high level interface that can interact with clinical tools or be operated by a non expert with consistent performance. This would certainly be necessary to enable real time adaptation while the patient is on the couch.

The first level of validation is the ability to faithfully reconstruction quantitatively accurate electron density. Through Polyquant and Polyquant–PolySKS, we have evaluated this in Chapter 3 and Chapter 4 with ground-truth numerical data and real measurements. The second level of validation was then to see if this gain in accuracy translated to an enhancement in dose calculation accuracy, which was evaluated in Chapter 5. From here, a third and final level of validation would be to evaluate the reduction of toxicity and cancer recurrence, when using this more accurate dose calculation, with associated tighter margins of uncertainty.

To minimise any harm that could be caused to patients, we believe this may be done retrospectively. One could collect CBCT data from patients receiving standard treatment protocols, and perform Polyquant reconstruction from this. One could then compare a dose plan according to these images, to the dose plan delivered to the patient from the planning CT. It may also be possible to acquire a dose measurement for validation from electron portal images of the exit dose. Firstly, any difference in dose between Polyquant and the original plan will likely indicate an error in the delivery, either under exposing the tumour, or over exposing healthy tissue. The long term effectiveness or toxicity of the radiation to the associated patients would then be strong evidence of a potential gain offered by our method.

Appendix A

Polyquant Fitting Parameters

Table A.1: Polyquant fitting parameters for ICRP 89 tissues [3] and Titanium, for relative electron density quantification, ρ_e.

energy (keV)	α_1	α_2	β_2	α_3	β_3
20	0.5827	7.4715	−6.9638	48.8113	−148.5237
25	0.4001	4.0995	−3.7396	26.6259	−80.8770
30	0.3081	2.3708	−2.0852	15.1340	−45.7901
35	0.2654	1.5644	−1.3132	9.7023	−29.1795
40	0.2400	1.0987	−0.8680	6.5471	−19.5248
45	0.2236	0.8208	−0.6037	4.6544	−13.7309
50	0.2122	0.6406	−0.4331	3.4254	−9.9690
55	0.2036	0.5184	−0.3183	2.5940	−7.4256
60	0.1970	0.4358	−0.2414	2.0349	−5.7172
65	0.1917	0.3789	−0.1892	1.6538	−4.5551
70	0.1871	0.3325	−0.1470	1.3433	−3.6082
75	0.1831	0.2959	−0.1140	1.1002	−2.8682
80	0.1795	0.2683	−0.0898	0.9215	−2.3265
85	0.1762	0.2470	−0.0715	0.7853	−1.9150
90	0.1733	0.2293	−0.0566	0.6724	−1.5738
95	0.1707	0.2151	−0.0449	0.5836	−1.3068
100	0.1683	0.2044	−0.0365	0.5202	−1.1180
105	0.1660	0.1957	−0.0301	0.4720	−0.9762
110	0.1638	0.1879	−0.0244	0.4282	−0.8473
115	0.1617	0.1808	−0.0193	0.3887	−0.7310
120	0.1597	0.1745	−0.0149	0.3532	−0.6269

Appendix B

Gate Simulation of CBCT System

The following script is for the simulation of a CBCT acquisition of the blabber region. This should be firstly split and ran in parallel.

```
/vis/disable
/control/execute mac/verbose.mac
/gate/geometry/setMaterialDatabase data/ICRP_HUmaterials.db
# World
/gate/world/setMaterial  Vacuum
/gate/world/geometry/setXLength  5 m
/gate/world/geometry/setYLength  5 m
/gate/world/geometry/setZLength  5 m
# Physics list
/gate/physics/addPhysicsList emlivermore
# Beam
/gate/source/addSource  mybeam gps
/gate/source/mybeam/gps/particle  gamma
/gate/source/mybeam/gps/ene/type  User
/gate/source/mybeam/gps/hist/type  energy
/gate/source/mybeam/gps/pos/type  UserFluenceImage
/gate/source/mybeam/gps/pos/setImage  data/source_prob_half_flip.mhd
/gate/source/mybeam/gps/centre  -2500 -357.104 -147.44 mm
/gate/source/mybeam/gps/ang/type  focused
/gate/source/mybeam/gps/ang/focuspoint  -1000 0 0 mm
/gate/source/mybeam/gps/pos/rot1  0 0 1
/control/execute  data/varian_spectrum.mac
# Insert phantom
/gate/world/daughters/name  patient
/gate/world/daughters/insert  ImageNestedParametrisedVolume
/gate/patient/geometry/setHUToMaterialFile  data/ICRP_HU2mat.txt
/gate/patient/geometry/setImage  data/icrp_blad_crop.mhd
/gate/patient/moves/insert  genericMove
/gate/patient/genericMove/setPlacementsFilename  data/rotation.placements
/gate/patient/vis/forceWireframe
# Detection plane
/gate/world/daughters/name  DetectorPlane
/gate/world/daughters/insert  box
/gate/DetectorPlane/geometry/setXLength  1 nm
/gate/DetectorPlane/geometry/setYLength  397.312 mm
/gate/DetectorPlane/geometry/setZLength  297.984 mm
/gate/DetectorPlane/placement/setTranslation  500 160 0 mm
/gate/DetectorPlane/setMaterial  Vacuum
/gate/DetectorPlane/vis/setVisible  1
/gate/DetectorPlane/vis/setColor  yellow
# Collimator
/gate/geometry/setMaterialDatabase  data/GateMaterials.db
/gate/world/daughters/name  collimator
/gate/world/daughters/insert  box
/gate/collimator/geometry/setXLength  2 mm
/gate/collimator/geometry/setYLength  450 mm
/gate/collimator/geometry/setZLength  350 mm
/gate/collimator/placement/setTranslation  498.5 160 0 mm
/gate/collimator/setMaterial  Aluminium
/gate/collimator/vis/setVisible  0
/gate/collimator/vis/setColor  blue
/gate/collimator/daughters/name  strip
/gate/collimator/daughters/insert  box
/gate/strip/geometry/setXLength  2 mm
/gate/strip/geometry/setYLength  400 mm
```

```
/gate/strip/geometry/setZLength  0.05 mm
/gate/strip/setMaterial  Lead
/gate/strip/vis/setVisible  1
/gate/strip/vis/setColor  green
/gate/strip/repeaters/insert  genericRepeater
/gate/strip/genericRepeater/setPlacementsFilename  data/collimator.repeat
/gate/strip/genericRepeater/useRelativeTranslation  0
# Recorded detector information
/gate/actor/addActor FluenceActor  DetectorFluence
/gate/actor/DetectorFluence/save  out_blad/detector.mhd
/gate/actor/DetectorFluence/attachTo  DetectorPlane
/gate/actor/DetectorFluence/stepHitType  pre
/gate/actor/DetectorFluence/setResolution  1 128 96
/gate/actor/DetectorFluence/enableScatter  true
/gate/actor/DetectorFluence/enableUncertainty  false
/gate/actor/DetectorFluence/responseDetectorFilename  data/responseCsI.txt
# Save particles stats
/gate/actor/addActor  SimulationStatisticActor stat
/gate/actor/stat/saveEveryNSeconds  20
/gate/actor/stat/save  out_blad/stats.txt
/gate/run/initialize
/gate/random/setEngineName  MersenneTwister
/gate/random/setEngineSeed  123456
/gate/application/setTimeSlice  6.25 s
/gate/application/setTimeStart  0 s
/gate/application/setTimeStop  1000 s
/gate/application/setTotalNumberOfPrimaries  1e11
/gate/application/start
exit
```

Bibliography

[1] A. J. Grossberg, A. S. R. Mohamed, S. Chamchod, W. Bennett, K. Smith, T. Nolan, M. Kantor, T. Browne, D. I. Rosenthal, and C. D. Fuller. (P055) TCIA Imaging Database for Head and Neck Squamous Cell Carcinoma Patients Treated With Radiotherapy. *Int. J. Radiat. Oncol.*, 98(2):E29–E30, jun 2017.

[2] A. Grossberg, A. Mohamed, H. Elhalawani, W. Bennett, K. Smith, T. Nolan, S. Chamchod, M. Kantor, T. Browne, K. Hutcheson, G. Gunn, A. Garden, S. Frank, D. Rosenthal, J. Freymann, and C. Fuller. Data from Head and Neck Cancer CT Atlas, 2017.

[3] ICRP Publication 89. Basic Anatomical and Physiological Data for Use in Radiological Protection Reference Values. *Ann. ICRP*, 32 (3-4), 2002.

[4] J. Yang, G. Sharp, H. Veeraraghavan, W. van Elmpt, A. Dekker, T. Lustberg, and M. Gooding. Data from Lung CT Segmentation Challenge, 2017.

[5] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging*, 26(6):1045–1057, dec 2013.

[6] W. H. Nailon, A. T. Redpath, and D. B. McLaren. Texture analysis of 3D bladder cancer CT images for improving radiotherapy planning. In *2008 5th IEEE Int. Symp. Biomed. Imaging From Nano to Macro*, pages 652–655. IEEE, may 2008.

[7] G. T. Herman and L. B. Meyer. Algebraic reconstruction techniques can be made computationally efficient (positron emission tomography application). *IEEE Trans. Med. Imaging*, 12(3):600–609, 1993.

[8] M. Sun and J. M. Star-Lack. Improved scatter correction using adaptive scatter kernel superposition. *Phys. Med. Biol.*, 55(22):6695–6720, nov 2010.

[9] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am. A*, 1(6):612, 1984.

[10] A. Beck and M. Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.*, 2(1):183–202, jan 2009.

[11] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Trans. Med. Imaging*, 20(10):999–1008, 2001.

[12] G.-H. Chen, J. Tang, and S. Leng. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Med. Phys.*, 35(2):660–663, 2008.

[13] Cancer Research UK. Cancer Statistics for the UK, 2018.

[14] National Cancer Registration & Analysis Service and Cancer Research UK. Chemotherapy, Radiotherapy and Tumour Resections in England: 2013-2014. *NCRAS*, 2017.

[15] A. Barrett, S. Morris, J. Dobbs, and T. Roques. *Practical Radiotherapy Planning Fourth Edition*. CRC Press, 2009.

[16] ICRU. Prescribing, Recording and Reporting Photon Beam Therapy. Technical report, 1993.

[17] ICRU. Prescribing, Recording and Reporting Photon Beam Therapy (Supplement to ICRU Report 50). Technical report, 1998.

[18] H. K. Tuy. An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math.*, 43(3):546–552, 1983.

[19] B. B. Thapa and J. A. Molloy. Feasibility of an image planning system for kilovoltage image-guided radiation therapy. *Med. Phys.*, 40(6Part1):061703, may 2013.

[20] J. Shepherd. Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review. *Polish J. Radiol.*, 79:181–193, 2014.

[21] Y.-K. Park, G. C. Sharp, J. Phillips, and B. A. Winey. Proton dose calculation on scatter-corrected CBCT image: Feasibility study for adaptive proton therapy. *Med. Phys.*, 42(8):4449–4459, 2015.

[22] P. Ochs, Y. Chen, T. Brox, and T. Pock. iPiano: Inertial Proximal Algorithm for

Non-Convex Optimization. apr 2014.

[23] D. Kim, S. Ramani, and J. A. Fessler. Combining Ordered Subsets and Momentum for Accelerated X-Ray CT Image Reconstruction. *IEEE Trans. Med. Imaging*, 34(1):167–178, jan 2015.

[24] A. Lomax, L. K. Folkes, and P. O'Neill. Biological Consequences of Radiation-induced DNA Damage: Relevance to Radiotherapy. *Clin. Oncol.*, 25(10):578–585, oct 2013.

[25] Brookhaven National Laboratory. Bragg Curves and Peaks, 2018.

[26] S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt, T. Frisson, L. Grevillot, L. Guigues, L. Maigne, C. Morel, Y. Perrot, N. Rehfeld, D. Sarrut, D. R. Schaart, S. Stute, U. Pietrzyk, D. Visvikis, N. Zahra, and I. Buvat. GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. *Phys. Med. Biol.*, 56(4):881–901, 2011.

[27] E. B. Podgorsak. *Radiation Physics for Medical Physicists*. Biological and Medical Physics, Biomedical Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[28] W. D. Newhauser and R. Zhang. The physics of proton therapy. *Phys. Med. Biol.*, 60(8):R155–R209, apr 2015.

[29] P. Wohlfahrt, C. Möhler, C. Richter, and S. Greilich. Evaluation of Stopping-Power Prediction by Dual- and Single-Energy Computed Tomography in an Anthropomorphic Ground-Truth Phantom. *Int. J. Radiat. Oncol.*, 100(1):244–253, jan 2018.

[30] U. Schneider, E. Pedroni, and A. Lomax. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Phys. Med. Biol.*, 41(1):111–124, 1996.

[31] J. Baró, J. Sempau, J. M. Fernández-Varea, and F. Salvat. PENELOPE: An algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms*, 100(1):31–46, may 1995.

[32] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau,

L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J.J. Gómez Cadenas, I. González, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F.W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M.G. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J.P. Wellisch, T. Wenaus, D.C. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche. Geant4a simulation toolkit. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, 506(3):250–303, jul 2003.

[33] M. Rodriguez, J. Sempau, and L. Brualla. PRIMO: A graphical environment for the Monte Carlo simulation of Varian and Elekta linacs. *Strahlentherapie und Onkol.*, 189(10):881–886, oct 2013.

[34] J. Sievinen, W. Ulmer, and W. Kaissl. AAA Photon Dose Calculation Model in Eclipse. Technical report, 2005.

[35] DKFZ. matRad, 2016. http://e0404.github.io/matRad/.

[36] I. I. Rosen, R. G. Lane, S. M. Morrill, and J. A. Belli. Treatment plan optimization using linear programming. *Med. Phys.*, 18(2):141–152, mar 1991.

[37] T. S. Curry, J. E. Dowdey, and R. C. Murry. *Christensen's Physics of Diagnostic Radiology*. 1990.

[38] J. M. Rini, A. Horowitz, S. Balter, and R. C. Watson. A Comparison of Tungsten

and Molybdenum as Target Material for Mammographic X-Ray Tubes. *Radiology*, 106(3):657–661, mar 1973.

[39] D. F. Jackson and D. J. Hawkes. X-ray attenuation coefficients of elements and mixtures. *Phys. Rep.*, 70(3):169–233, apr 1981.

[40] O. Klein and Y. Nishina. Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. *Zeitschrift für Phys.*, 52(11-12):853–868, nov 1929.

[41] ICRP Publication 110. Adult Reference Computational Phantoms. *Ann. ICRP*, 39 (2), 2009.

[42] R. E. Alvarez and A. Macovski. Energy-selective reconstructions in X-ray computerized tomography. *Phys. Med. Biol.*, 21:733–744, 1976.

[43] R. A. Rutherford, B. R. Pullan, and I. Isherwood. Measurement of effective atomic number and electron density using an EMI scanner. *Neuroradiology*, 11(1):15–21, 1976.

[44] J. Weber and D. J. van den Berge. The effective atomic number and the calculation of the composition of phantom materials. *Br. J. Radiol.*, 42(497):378–383, may 1969.

[45] P. M. Joseph and R. D. Spital. A method for correcting bone induced artifacts in computed tomography scanners., 1978.

[46] I. A. Elbakri and J. A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Med. Imaging*, 21(2):89–99, feb 2002.

[47] I. A. Elbakri and J. A. Fessler. Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation. *Phys. Med. Biol.*, 48(15):2453–2477, aug 2003.

[48] Y. Meng, C. C. Shaw, X. Liu, M. C. Altunbas, T. Wang, L. Chen, S.-J. Tu, S. C. Kappadath, and C.-J. Lai. Comparison of two detector systems for cone beam CT small animal imaging: a preliminary study. page 614245, mar 2006.

[49] Q. Cao, A. Sisniega, M. Brehler, J. W. Stayman, J. Yorkston, J. H. Siewerdsen, and W. Zbijewski. Modeling and evaluation of a high-resolution CMOS detector for cone-beam CT of the extremities. *Med. Phys.*, 45(1):114–130, jan 2018.

[50] X. Jia, H. Yan, L. Cerviño, M. Folkerts, and S. B. Jiang. A GPU tool for efficient, accurate, and realistic simulation of cone beam CT projections. *Med. Phys.*, 39(12):7368–7378, nov 2012.

[51] M. Freed, S. Park, and A. Badano. A fast, angle-dependent, analytical model of CsI detector response for optimization of 3D x-ray breast imaging systems. *Med. Phys.*, 37(6Part1):2593–2605, may 2010.

[52] C. Richmond. Sir Godfrey Hounsfield. *BMJ*, 329(7467):687.1, sep 2004.

[53] E. C. Beckmann. CT scanning the early days. *Br. J. Radiol.*, 79(937):5–8, jan 2006.

[54] Centre of Evidence Based Purchasing. Buyers' guide: Multi-slice CT scanners. Technical report, NHS, 2009.

[55] J. A. Fessler. Fundamentals of CT Reconstruction in 2D and 3D. In *Compr. Biomed. Phys.*, pages 263–295. Elsevier, 2014.

[56] D. A. Jaffray and J. H. Siewerdsen. Cone-beam computed tomography with a flat-panel imager: Initial performance characterization. *Med. Phys.*, 27(6):1311–1323, jun 2000.

[57] G. Wang. X-ray micro-CT with a displaced detector array. *Med. Phys.*, 29(7):1634, 2002.

[58] S. Schafer, J. W. Stayman, W. Zbijewski, C. Schmidgunst, G. Kleinszig, and J. H. Siewerdsen. Antiscatter grids in mobile C-arm cone-beam CT: Effect on image quality and dose. *Med. Phys.*, 39(1):153–159, dec 2011.

[59] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman. Modelling the physics in the iterative reconstruction for transmission computed tomography. *Phys. Med. Biol.*, 58(12):R63–96, jun 2013.

[60] Z. Chang, R. Zhang, J.-B. Thibault, K. Sauer, and C. Bouman. Statistical x-ray computed tomography imaging from photon-starved measurements. *SPIE Comput. Imaging*, 9020:90200G, 2014.

[61] X. Pan, E. Y. Sidky, and M. Vannier. Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse Probl.*, 25(12):123009, dec 2009.

[62] B. D. Smith. Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods. *IEEE transactions on medical imaging*, 4(1):14–25, 1985.

[63] P. Grangeat. Mathematical framework of cone beam 3D reconstruction via the first derivative of the radon transform. pages 66–97. 1991.

[64] S. Bartolac, R. Clackdoyle, F. Noo, J. Siewerdsen, D. Moseley, and D. Jaffray. A local shift-variant Fourier model and experimental validation of circular cone-beam computed tomography artifacts. *Medical Physics*, 36(2):500–512, jan 2009.

[65] A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, 2001.

[66] H. Hui. An improved cone-beam reconstruction algorithm for the circular orbit. *Scanning*, 18:572–581, 1996.

[67] F. Arcadu, M. Stampanoni, and F. Marone. On the crucial impact of the coupling projector-backprojector in iterative tomographic reconstruction. dec 2016.

[68] R. M. Lewitt. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. *J. Opt. Soc. Am. A.*, 7(10):1834–46, oct 1990.

[69] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49(11):2463–75, jun 2004.

[70] Y. Long, J. A. Fessler, and J. M. Balter. 3D Forward and Back-Projection for X-Ray CT Using Separable Footprints. *IEEE Trans. Med. Imaging*, 29(11):1839–1850, nov 2010.

[71] A. Wirgin. The inverse crime. jan 2004.

[72] J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.*, 198(2):493–504, jan 2007.

[73] K. Jensen, H. K. Andersen, Ö. Smedby, B. H. Østerås, A. Aarsnes, A. Tingberg, E. Fosse, and A. C. Martinsen. Quantitative Measurements Versus Receiver Operating Characteristics and Visual Grading Regression in CT Images Reconstructed with Iterative Reconstruction. *Acad. Radiol.*, 25(4):509–518, apr 2018.

[74] ICRP Publication 103. The 2007 Recommendations of the International Commission on Radiological Protection. *Ann. ICRP*, 2007.

[75] United Nations Scientific Committee on the Effects of Atomic Radiation. UNSCEAR 2008 Report to the General Assembly. Technical report, 2008.

[76] P. Alaei and E. Spezi. Imaging dose from cone beam computed tomography in radiation therapy. *Phys. Medica*, 31(7):647–658, nov 2015.

[77] B. Meng, L. Xing, B. Han, A. Koong, D. Chang, J. Cheng, and R. Li. Cone beam CT imaging with limited angle of projections and prior knowledge for volumetric verification of non-coplanar beam radiation therapy: a proof of concept study. *Phys. Med. Biol.*, 58(21):7777–7789, nov 2013.

[78] Z. Zhang, Xiao H., Junguo B., Joseph J. M., E. Y. Sidky, and X. Pan. Initial experience in image reconstruction from limited-angle C-arm CBCT data. In *2011 IEEE Nucl. Sci. Symp. Conf. Rec.*, pages 3977–3979. IEEE, oct 2011.

[79] C. Mory, V. Auvray, B. Zhang, M. Grass, D. Schäfer, S. J. Chen, J. D. Carroll, S. Rit, F. Peyrin, P. Douek, and L. Boussel. Cardiac C-arm computed tomography using a 3D + time ROI reconstruction method with spatial and temporal regularization. *Med. Phys.*, 41(2):021903, jan 2014.

[80] W. M. Thompson, W. R. B. Lionheart, E. J. Morton, M. Cunningham, and R. D. Luggar. High speed imaging of dynamic processes with a switched source x-ray CT system. *Measurement Science and Technology*, 26(5):055401, may 2015.

[81] S. Coban, W. R. B. Lionheart, P. J. Withers, and S. A. McDonald. Sophia B Coban, William RB Lionheart, Philip J Withers, Samuel A McDonald. In *13th Int. Meet. Fully Three-Dimensional Image Reconstr. Radiol. Nucl. Med.*, Newport, Rhode Island, USA, 2015.

[82] J. Nocedal. Updating quasi-Newton matrices with limited storage. *Math. Comput.*, 35(151):773–773, sep 1980.

[83] B. De Man, S. Basu, J.-B. Thibault, J. Hsieh, J. A. Fessler, C. Bouman, and K. Sauer. A Study of Different Minimization Approaches for Iterative Reconstruction in X-Ray CT. In *IEEE Nucl. Sci. Symp. Conf. Rec. 2005*, volume 5, pages 2708–2710. IEEE, 2005.

[84] A. D. Klose and A. H. Hielscher. Quasi-Newton methods in optical tomographic image reconstruction. *Inverse Probl.*, 19(2):387–409, apr 2003.

[85] N. H. Clinthorne, T.-S. Pan, P.-C. Chiao, W. L. Rogers, and J. A. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans. Med. Imaging*, 12(1):78–83, mar 1993.

[86] H. Erdogan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Trans. Med. Imaging*, 18(9):801–814, 1999.

[87] H. Dang, J. W. Stayman, A. Sisniega, J. Xu, W. Zbijewski, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen. Statistical reconstruction for cone-beam CT with a post-artifact-correction noise model: application to high-quality head imaging. *Phys. Med. Biol.*, 60(16):6153–6175, aug 2015.

[88] Y. Chen, D. Gao, C. Nie, L. Luo, W. Chen, X. Yin, and Y. Lin. Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior. *Comput. Med. Imaging Graph.*, 33(7):495–500, oct 2009.

[89] P. J. Huber. Robust Estimation of a Location Parameter. *Ann. Math. Stat.*, 35(1):73–101, mar 1964.

[90] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, 6(2):298–311, 1997.

[91] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. 1st Int. Conf. Image Process.*, volume 2, pages 168–172. IEEE Comput. Soc. Press, 1994.

[92] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, feb 2006.

[93] D. L. Donoho. Compressed Sensing. 52(4):1289–1306, 2006.

[94] J. S. Jorgensen and E. Y. Sidky. How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 373(2043):20140387–20140387, may 2015.

[95] A. Haar. Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.*, 69(3):331–371, sep 1910.

[96] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, jan 1992.

[97] M. Elad and M. Aharon. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, dec 2006.

[98] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.*, 60(1-4):259–268, nov 1992.

[99] J. S. Jørgensen, S. B. Coban, W. R. B. Lionheart, S. A. McDonald, and P. J. Withers. SparseBeads data: benchmarking sparsity-regularized computed tomography. *Meas. Sci. Technol.*, 28(12):124005, dec 2017.

[100] E. Y. Sidky, C. M. Kao, and X. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *J. Xray. Sci. Technol.*, pages 1–30, 2006.

[101] C. R. Vogel and M. E. Oman. Iterative Methods for Total Variation Denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, jan 1996.

[102] T. Goldstein, C. Studer, and R. Baraniuk. A Field Guide to Forward-Backward Splitting with a FASTA Implementation. nov 2014.

[103] A. Chambolle. An algorithm for total variation minimizations and applications. *J. Math. Imaging Vis.*, 10(1-2):89–97, jan 2004.

[104] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, 2009.

[105] S. Anthoine, J.-F. Aujol, Y. Boursier, and C. Melot. On the efficiency of proximal methods in CBCT and PET. In *2011 18th IEEE Int. Conf. Image Process.*, pages 1365–1368. IEEE, sep 2011.

[106] H. Zhang, J. Ma, Z. Bian, D. Zeng, Q. Feng, and W. Chen. High quality 4D cone-beam CT reconstruction using motion-compensated total variation regularization.

Phys. Med. Biol., 62(8):3313–3329, apr 2017.

[107] J. W. Stayman, H. Dang, Y. Ding, and J. H. Siewerdsen. PIRPLE: a penalized-likelihood framework for incorporation of prior images in CT reconstruction. *Phys. Med. Biol.*, 58(21):7563–7582, nov 2013.

[108] H. Dang, A. S. Wang, M. S. Sussman, J. H. Siewerdsen, and J. W. Stayman. dPIRPLE: a joint estimation framework for deformable registration and penalized-likelihood CT image reconstruction using prior images. *Phys. Med. Biol.*, 59(17):4799–826, sep 2014.

[109] A. Pourmorteza, H. Dang, J. H. Siewerdsen, and J. W. Stayman. Reconstruction of difference in sequential CT studies using penalized likelihood estimation. *Phys. Med. Biol.*, 61(5):1986–2002, mar 2016.

[110] K. Choi, J. Wang, L. Zhu, T.-Suk. Suh, S. Boyd, and L. Xing. Compressed sensing based cone-beam computed tomography reconstruction with a first-order method. *Med. Phys.*, 37(9):5113–5125, aug 2010.

[111] E. Chouzenoux, J. C. Pesquet, and A. Repetti. Variable Metric Forward-Backward Algorithm for Minimizing the Sum of a Differentiable Function and a Convex Function. *J. Optim. Theory Appl.*, 162(1):107–132, 2014.

[112] S. Ramani and J. A. Fessler. A Splitting-Based Iterative Algorithm for Accelerated Statistical X-Ray CT Reconstruction. *IEEE Trans. Med. Imaging*, 31(3):677–688, mar 2012.

[113] I. Goodfellow, Y. Bengio, and A. Courville. Optimization for Training Deep Models. In *Deep Learn.*, pages 271–325. 2016.

[114] H. Erdogan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, nov 1999.

[115] D. S. Lalush and B. M. Tsui. Performance of ordered-subset reconstruction algorithms under conditions of extreme attenuation and truncation in myocardial SPECT. *J. Nucl. Med.*, 41(4):737–44, apr 2000.

[116] S. Ahn and J. A. Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imaging*, 22(5):613–626, may 2003.

[117] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.*, 4(5):1–17, jan 1964.

[118] Y. Nesterov. A Method for Solving a Convex Programming Problem with Convergence Rate O(1/K2). *Sov. Math. Dokl.*, 27:372–367, 1983.

[119] A. S. Wang, J. W. Stayman, Y. Otake, S. Vogt, G. Kleinszig, and J. H. Siewerdsen. Accelerated statistical reconstruction for C-arm cone-beam CT using Nesterov's method. *Med. Phys.*, 42(5):2699–2708, may 2015.

[120] P. Després and X. Jia. A review of GPU-based medical image reconstruction. *Phys. Medica*, 42:76–92, oct 2017.

[121] H. Scherl, M. Kowarschik, H. G. Hofmann, B. Keck, and J. Hornegger. Evaluation of state-of-the-art hardware architectures for fast cone-beam CT reconstruction. *Parallel Comput.*, 38(3):111–124, mar 2012.

[122] M. Leeser, S. Mukherjee, and J. Brock. Fast reconstruction of 3D volumes from 2D CT projection data with GPUs. *BMC Res. Notes*, 7(1):582, 2014.

[123] F. Xu, W. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard, and K. Mueller. On the efficiency of iterative ordered subset reconstruction algorithms for acceleration on GPUs. *Comput. Methods Programs Biomed.*, 98(3):261–270, jun 2010.

[124] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang. GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation. *Med. Phys.*, 37(4):1757–1760, mar 2010.

[125] J. F. Barrett and N. Keat. Artifacts in CT: Recognition and Avoidance. *RadioGraphics*, 24(6):1679–1691, nov 2004.

[126] W. D. McDavid, R. G. Waggener, W. H. Payne, and M. J. Dennis. Correction for spectral artifacts in cross-sectional reconstruction from x rays. *Medical Physics*, 4(1):54–57, jan 1977.

[127] G. T. Herman. Correction for beam hardening in computed tomography. *Physics in Medicine and Biology*, 24(1):008, jan 1979.

[128] E.-P. Rührnschopf and K. Klingenbeck. A general framework and review of scatter correction methods in cone beam CT. Part 2: Scatter estimation approaches. *Med.*

Phys., 38(9):5186–5199, aug 2011.

[129] E.-P. Rührnschopf and K. Klingenbeck. A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. Part 1: Scatter compensation approaches. *Med. Phys.*, 38(7):4296–4311, jun 2011.

[130] L. Zhu, N. R. Bennett, and R. Fahrig. Scatter Correction Method for X-Ray CT Using Primary Modulation: Theory and Preliminary Results. *IEEE Trans. Med. Imaging*, 25(12):1573–1587, dec 2006.

[131] Y. Xu, T. Bai, H. Yan, L. Ouyang, A. Pompos, J. Wang, L. Zhou, S. B. Jiang, and X. Jia. A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy. *Phys. Med. Biol.*, 60(9):3567–3587, may 2015.

[132] G. Poludniowski, P. M. Evans, V. N. Hansen, and S Webb. An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT. *Phys. Med. Biol.*, 54(12):3847–3864, 2009.

[133] W. Zbijewski and F. J. Beekman. Efficient Monte Carlo based scatter artifact reduction in cone-beam micro-CT. *IEEE Trans. Med. Imaging*, 25(7):817–827, jul 2006.

[134] A. Sisniega, W. Zbijewski, J. Xu, H. Dang, J. W. Stayman, J. Yorkston, N. Aygun, V. Koliatsos, and J. H. Siewerdsen. High-fidelity artifact correction for cone-beam CT imaging of the brain. *Phys. Med. Biol.*, 60(4):1415–1439, feb 2015.

[135] J.-S. Lee and J.-C. Chen. A Single Scatter Model for X-ray CT Energy Spectrum Estimation and Polychromatic Reconstruction. *IEEE Trans. Med. Imaging*, 34(6):1403–1413, jun 2015.

[136] B. Ohnesorge, T. Flohr, and K. Klingenbeck-Regn. Efficient object scatter correction algorithm for third and fourth generation CT scanners. *Eur. Radiol.*, 9(3):563–569, mar 1999.

[137] J. H. Siewerdsen and D. A. Jaffray. Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter. *Med. Phys.*, 28(2):220, 2001.

[138] J. Y. Lo, C. E. Floyd, J. A. Baker, and C. E. Ravin. An artificial neural network for estimating scatter exposures in portable chest radiography. *Med. Phys.*, 20(4):965–

973, jul 1993.

[139] J. Maier, S. Sawall, M. Kachelriess, and Y. Berker. Deep scatter estimation (DSE): feasibility of using a deep convolutional neural network for real-time x-ray scatter prediction in cone-beam CT. In Guang-Hong Chen, Joseph Y. Lo, and Taly Gilat Schmidt, editors, *Med. Imaging 2018 Phys. Med. Imaging*, page 56. SPIE, mar 2018.

[140] J. H. Mason, W. H. Nailon, and M. E. Davies. A Monte Carlo Framework for Low Dose CT Reconstruction Testing. In *Simulation and Synthesis in Medical Imaging*, pages 79–88. Springer, 2017.

[141] E. B. Saloman, J. H. Hubbell, and J. H. Scofield. X-ray attenuation cross sections for energies 100 eV to 100 keV and elements Z = 1 to Z = 92. *At. Data Nucl. Data Tables*, 38(1):1–196, jan 1988.

[142] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, jan 1998.

[143] G. Poludniowski, G. Landry, F. DeBlois, P. M. Evans, and F. Verhaegen. SpekCalc: a program to calculate photon spectra from tungsten anode x-ray tubes. *Phys. Med. Biol.*, 54(19):N433–N438, oct 2009.

[144] J. J. Battista, W. D. Rider, and J. Van Dyk. Computed tomography for radiotherapy planning. *Int. J. Radiat. Oncol.*, 6(1):99–107, jan 1980.

[145] D. Marshall, O. Johnell, and H. Wedel. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ*, 312(7041):1254–1259, may 1996.

[146] T. R. C. Johnson. Dual-Energy CT: General Principles. *Am. J. Roentgenol.*, 199(5_supplement):S3–S8, nov 2012.

[147] C. Constantinou, J. C. Harrington, and L. A. DeWerd. An electron density calibration phantom for CT-based treatment planning computers. *Med. Phys.*, 19(2):325–327, mar 1992.

[148] J. H. Mason, A. Perelli, W. H. Nailon, and M. E. Davies. Quantitative Electron Density CT Imaging for Radiotherapy Planning. In *Med. Image Underst. Anal.*, pages 297–308. Springer, 2017.

[149] J. H. Mason, A. Perelli, W. H. Nailon, and M. E. Davies. Polyquant CT: direct electron and mass density reconstruction from a single polyenergetic source. *Phys. Med. Biol.*, oct 2017.

[150] A. Ritter and N. Mistry. *DirectDensity: Technical principles and implications for radiotherapy (white paper).* Siemens Healthineers, 2016.

[151] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Process.*, 2(3):296–310, jul 1993.

[152] I. Daubechies, M. Fornasier, and I. Loris. Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints. *J. Fourier Anal. Appl.*, 14(5-6):764–792, dec 2008.

[153] J. A. Fessler. Image Reconstruction Toobox, 2018.

[154] L. Fu, S. M. Kim, A. M. Alessio, P. E. Kinahan, and B. De Man. Comparison between pre-log and post-log statistical models in low-dose CT iterative reconstruction. In *2014 IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, pages 1–10. IEEE, nov 2014.

[155] J. S. Maltz, B. Gangadharan, S. Bose, D. H. Hristov, B. A. Faddegon, A. Paidi, and A. R. Bani-Hashemi. Algorithm for X-ray Scatter, Beam-Hardening, and Beam Profile Correction in Diagnostic (Kilovoltage) and Treatment (Megavoltage) Cone Beam CT. *IEEE Trans. Med. Imaging*, 27(12):1791–1810, dec 2008.

[156] M. Sun, T. Nagy, G. Virshup, L. Partain, M. Oelhafen, and J. Star-Lack. Correction for patient table-induced scattered radiation in cone-beam computed tomography (CBCT). *Med. Phys.*, 38(4):2058–2073, mar 2011.

[157] M. Sun, A. Maslowski, I. Davis, T. Wareing, G. Failla, and J. Star-Lack. Rapid scatter estimation for CBCT using the Boltzmann transport equation. page 90330Z, mar 2014.

[158] J. H. Mason, A. Perelli, W. H. Nailon, and M. E. Davies. Can Planning Images Reduce Scatter in Follow-Up Cone-Beam CT? In *Medical Image Understanding and Analysis*, pages 629–640. Springer, 2017.

[159] C. Davisson and R. Evans. Gamma-Ray Absorption Coefficients. *Rev. Mod. Phys.*, 24(2):79–107, apr 1952.

[160] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack. Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - Part I: Core algorithms and validation. *Med. Phys.*, apr 2018.

[161] A. Wang, A. Maslowski, P. Messmer, M. Lehmann, A. Strzelecki, E. Yu, P. Paysan, M. Brehm, P. Munro, J. Star-Lack, and D. Seghers. Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - Part II: System modeling, scatter correction, and optimization. *Med. Phys.*, mar 2018.

[162] H. Yan, X. Wang, F. Shi, T. Bai, M. Folkerts, L. Cervino, S. B. Jiang, and X. Jia. Towards the clinical implementation of iterative low-dose cone-beam CT reconstruction in image-guided radiation therapy: Cone/ring artifact correction and multiple GPU implementation. *Med. Phys.*, 41(11):111912, oct 2014.

[163] T. Niu, A. Al-Basheer, and L. Zhu. Quantitative cone-beam CT imaging in radiation therapy using planning CT as a prior: First patient studies. *Med. Phys.*, 39(4):1991–2000, mar 2012.

[164] O. Weistrand and S. Svensson. The ANACONDA algorithm for deformable image registration in radiotherapy. *Med. Phys.*, 42(1):40–53, dec 2014.

[165] S. M. Bentzen, L. S. Constine, J. O. Deasy, A. Eisbruch, A. Jackson, L. B. Marks, R. K. Ten Haken, and E. D. Yorke. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. *Int. J. Radiat. Oncol.*, 76(3):S3–S9, mar 2010.

[166] L. B. Marks, E. D. Yorke, A. Jackson, R. K. Ten Haken, L. S. Constine, A. Eisbruch, S. M. Bentzen, J. Nam, and J. O. Deasy. Use of Normal Tissue Complication Probability Models in the Clinic. *Int. J. Radiat. Oncol.*, 76(3):S10–S19, mar 2010.

[167] M. A. Schmidt and G. S. Payne. Radiotherapy planning using MRI. *Phys. Med. Biol.*, 60(22):R323–R361, nov 2015.

[168] H. Raguet, J. Fadili, and G. Peyré. A Generalized Forward-Backward Splitting. *SIAM J. Imaging Sci.*, 6(3):1199–1226, jan 2013.

[169] J. Xu, A. Sisniega, W. Zbijewski, H. Dang, J. W. Stayman, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen. Modeling and design of a cone-beam CT head scanner using task-based imaging performance optimization. *Phys.*

Med. Biol., 61(8):3180–3207, apr 2016.

[170] S. Xu, A. Uneri, A. J. Khanna, J. H. Siewerdsen, and J. W. Stayman. Polyenergetic known-component CT reconstruction with unknown material compositions and unknown x-ray spectra. *Phys. Med. Biol.*, 62(8):3352–3374, apr 2017.

[171] N. M. Wink, C. Panknin, and T. D. Solberg. Phase versus amplitude sorting of 4D-CT data. *J. Appl. Clin. Med. Phys.*, 7(1):77–85, dec 2006.

[172] J. A. Fessler, I. A. Elbakri, P. Sukovic, and N. H. Clinthorne. Maximum-likelihood dual-energy tomographic image reconstruction. In Milan Sonka and J. Michael Fitzpatrick, editors, *SPIE Medical Imaging*, page 38, may 2002.